DMT of Multi-hop Cooperative Networks - Part I: Basic Results

K. Sreeram, S. Birenjith, and P. Vijay Kumar

***Abstract*— In this two-part paper, the DMT of cooperative multi-hop networks is examined. The focus is on single-source single-sink (ss-ss) multi-hop relay networks having slow-fading links and relays that potentially possess multiple antennas. In this first part, some basic results that help in determining the DMT of cooperative networks as well as in characterizing the two end-points of the DMT for arbitrary full-duplex networks is established. In the companion paper, two families of half-duplex networks are studied.**

The present paper examines the two end-points of the DMT of ss-ss networks. In particular, the maximum achievable diversity of arbitrary multi-terminal wireless networks is shown to be equal to the min-cut between the corresponding source and the sink. The maximum multiplexing gain (MMG) of arbitrary full-duplex ss-ss networks is shown to be equal to the min-cut rank, using a new connection to a deterministic network for which the capacity was recently found. This connection is operational in the sense that a capacity-achieving scheme for the deterministic network can be converted into a MMG-achieving scheme for the original network.

We also prove some basic results including a proof that the colored noise encountered in AF protocols for cooperative networks can be treated as white noise for DMT computations. We derive lower bounds for the DMT of triangular channel matrices, which are subsequently utilized to derive alternative, and often simpler proofs of several existing results. The DMT of a parallel channel with independent MIMO links is also computed here. As an application of these basic results, we prove that a linear tradeoff between maximum diversity and maximum multiplexing gain is achievable for arbitrary, ss-ss single-antenna, directed-acyclic networks equipped with full-duplex relays.

All protocols in this paper are explicit and rely only upon amplify-and-forward (AF) relaying. Explicit codes for all protocols introduced here are included in the companion paper.

I. Introduction

In fading relay networks, cooperative diversity provides a means of operating the network efficiently. While much of the work in the literature on cooperative diversity is based on two-hop networks, the attention here is on multi-hop networks.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India (Email: sreeramkannan@ece.iisc.ernet.in, biren@ece.iisc.ernet.in, vijayk@usc.edu). P. Vijay Kumar is on leave of absence from the University of Southern California, Los Angeles, USA.

This work was supported in part by NSF-ITR Grant CCR-0326628, in part by the DRDO-IISc Program on Advanced Mathematical Engineering and in part by Motorola's University Research Partnership Program with IISc.

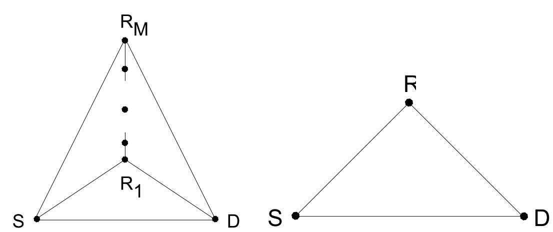
The material in this paper was presented in part at the 10th International Symposium on Wireless Personal Multimedia Communications, Jaipur, Dec. 2007, the Information Theory and Applications Workshop, San Diego, Jan. 2008, and at the IEEE International Symposium on Information Theory, Toronto, July 2008.

A. Prior Work

The concept of user cooperative diversity was introduced in [2]. Cooperative diversity protocols were first discussed in [3] for the two-hop, single-relay network (Fig.1(a)). Zheng and Tse [4] proposed the Diversity-Multiplexing gain Tradeoff (DMT) as a means of evaluating point-to-point, multiple-antenna schemes in the context of slow-fading channels.

1) Two-hop Networks: The DMT was also used as a tool to compare various protocols for half-duplex two-hop cooperative networks in [5], [6]. As noted in [9], the DMT is simple enough to be analytically tractable and powerful enough to compare different cooperative-relay-network protocols. For any network, an upper bound on the achievable DMT is given by the cut-set bound [9], [49]. A fundamental question in this area is whether the cut-set bound on DMT can be achieved. While this question has been studied extensively for the two-hop cooperative wireless system in Fig.1(a), the question still remains open even for this class of network (see [10], [12] for a detailed comparison of existing achievable regions).

In [5], the selection-decode-and-forward protocol is analyzed for an arbitrary number of relays, where the authors give upper and lower bounds on the DMT of the protocol. In these protocols, the relays and the source node participate for equal time instants and the maximum multiplexing gain r achieved is equal to 0.5.



(a) General two-hop relay network (b) Single relay network

Fig. 1. Two-hop cooperative relay networks

In [6], Azarian *et al.* analyze the class of Non Orthogonal, Amplify-and-Forward (NAF) protocols, introduced earlier by Nabar *et al.* in [7] and establish the improved DMT of the NAF protocol in comparison to the class of Orthogonal-Amplify-and-Forward (OAF) protocols considered in [5]. It has been shown in [10], that the DMT of the NAF protocol can be obtained via the OAF protocol as well using appropriate unequal slot lengths for source and relay transmissions.

The authors of [6] also introduce the Dynamic Decode-and-Forward (DDF) protocol wherein the time duration for which

the relays listen to the source depends on the source-relay channel gain. They show that for the single-relay case, the DMT of the DDF protocol achieves the cut-set bound (also known as the transmit-diversity bound) for $r \leq 0.5$, beyond which the DMT falls below the bound. An enhanced DDF protocol is proposed in [12] that improves upon DDF. However the DMT of this protocol also falls short of the transmit bound for $r \geq 0.5$.

Yang and Belfiore consider a class of protocols called Slotted-Amplify-And-Forward (SAF) protocols in [19] for the two-hop network with direct link, and show that these improve upon the performance of the NAF protocol [6] for the case of two relays. Under the assumption of relay isolation, the naive SAF scheme proposed in [19] is shown to achieve the cut-set bound. It is also conjectured in [19] that SAF protocol is optimal even when the relays are not isolated.

Yuksel and Erkip in [9] have considered the DMT of the DF and compress-and-forward (CF) protocols. They show that the CF protocol achieves the transmit-diversity bound for the case of a single relay. We note, however, that in the CF protocol, the relays are assumed to know all the fading coefficients in the system. The authors also translate cut-set upper bounds in [49] for mutual information into the DMT framework for a general multi-terminal network.

Jing and Hassibi [8] consider cooperative communication protocols for the two-hop network without a direct link between source and destination. They study protocols where the relay nodes apply a linear transformation to the received signal and analyze their BER performance. The authors consider the case when the source and the relays transmit for an equal number of channel uses and the relays perform a unitary transformation on the input symbols before transmitting it. Rao and Hassibi [28] consider two-hop half-duplex multi-antenna cooperative networks without direct link and provide an AF scheme and compute the DMT achieved by the scheme. Their scheme incurs a rate loss of a factor of two compared to the cut-set bound. In a parallel work [30], the DMT of the two-hop network without direct link is proved to be equal to the cut-set bound.

2) Multi-hop Networks: Yang and Belfiore in [18] consider AF protocols for a family of MIMO multi-hop networks (which are termed as multi-antenna layered networks in the current paper). They derive the optimal DMT for the Rayleigh-product channel which they prove is equal to the DMT of the AF protocol applied to this channel. They also propose AF protocols to achieve the optimal diversity of these multi-antenna layered networks.

Oggier and Hassibi [40] have proposed distributed space time codes for multi-antenna layered networks that achieve diversity gain equal to the minimum number of relay nodes among the hops. Recently, Vaze and Heath [41] have constructed distributed space time codes based on orthogonal designs that achieve the optimal diversity of the multi-antenna layered network with low decoding complexity. In [42], the same authors study the circumstances under which full diversity can be achieved without coding in a layered network in the presence of partial CSIT.

Borade, Zheng and Gallager in [27] consider AF schemes on a class of multi-hop layered networks where each layer has the same number of relays (termed as regular networks in the current paper). They show that AF strategies are optimal in terms of multiplexing gain. They also compute lower bounds on the DMT of the product Rayleigh channel.

3) Capacity: There has been a recent interest in determining approximations to the capacity of wireless networks. The pre-log coefficient of the capacity, termed as the degrees-of-freedom (DOF) of wireless multi-antenna networks is studied in [45] [46]. [1] The DOF for the N user interference channel was derived in [35], for the MIMO X networks in [36], [37] and the DOF of single-source single-sink (ss-ss) layered networks was obtained in [27].

In a different direction, the capacity of ss-ss and multi-cast deterministic wireless networks has been characterized in [31].

Intuition drawn from the deterministic wireless networks was used to identify capacity to within a constant for some example networks in [32]. Very recently, the capacity of single-antenna gaussian relay networks has been characterized to within a constant number of bits in [33]. This result also easily extends to give the approximate compound-channel capacity for full-duplex single-antenna networks. The results in [33] can also be used to show that for half duplex networks, under any fixed schedule of operation, the best possible rate can be achieved (to within a constant number of bits). However the determination of optimal schedules that achieve the maximum possible DMT remains open, which we solve for certain classes of networks in the companion paper.

In [34], given a wireline network code, a scheme for wireless gaussian relay channel is obtained where each relay computes linear transformations of its input signals and the achievable rate region for the scheme is characterized.

4) Codes: Cyclic Division Algebras (CDA) were first used to construct space-time codes in [20]. The notion of space-time codes having a non-vanishing determinant (NVD) was introduced in [11]. Subsequently, it was shown in [14] that CDA-based ST codes with NVD achieve the DMT of the Rayleigh-fading channel and minimal-delay codes with NVD were constructed for all n_t. From the results in [13], these codes are moreover, approximately universal, i.e., DMT optimal for every statistical characterization of the fading channel.

These codes were tailored to suit the structure of various static protocols for two-hop cooperation and proved to be DMT optimal for certain protocols in [10]. For the DDF protocol, DMT optimal codes were constructed for arbitrary number of relays with multiple antennas in [15]. For the specific case of single-relay single-antenna DDF channel, codes were constructed recently in [16], which are not only DMT optimal, but also have probability of error close to the outage probability. Codes for the multi-antenna two-hop network under the NAF protocol were presented in [17]. CDA-based ST codes construction for the rayleigh parallel channel were provided in [43], [44]. This construction was shown to

[1] The degrees-of-freedom is alternately referred to as maximum multiplexing gain in the literature, although the former is typically used for ergodic capacity characterizations and the latter is typically used in the context of outage characterization. This paper deals with the DMT, which is a outage characterization and for this reason, we use the term multiplexing gain.

be approximately universal for the class of MIMO parallel channels in [15]. In this paper, we present a DMT optimal code design for all proposed protocols based on the approximately universal codes in [14] and [15].

5) Other Work: Cooperative networks with asynchronous transmissions have also been studied in the literature [54], [55], [56]. However, we consider networks in which relays are synchronized. Codes for two-hop cooperative networks having low decoding complexity and full diversity are studied in [57], [56] and [58]. While decoding complexity is not the primary focus of the present paper, we do provide a successive-interference-cancellation technique to reduce the code length and thereby, the complexity.

B. Setting and Channel Model

1) Network Representation by a Graph: Unless otherwise stated, all networks considered possess a single source and a single sink and we will apply the abbreviation ss-ss to denote these networks. Any wireless network can be associated with a directed graph, with vertices representing nodes in the network and edges representing connectivity between nodes. If an edge is bidirectional, we will represent it by two edges, one pointing in either direction. An edge in a directed graph is said to be *live* at a particular time instant if the node at the head of the edge is transmitting at that instant. An edge in a directed graph is said to be *active* at a particular time instant if the node at the head of the edge is transmitting and the tail of the edge is receiving at that instant.

A wireless network is characterized by broadcast and interference constraints. Under the *broadcast* constraint, all edges connected to a transmitting node are simultaneously live and transmit the same information. Under the *interference* constraint, the symbol received by a receiving end is equal to the sum of the symbols transmitted on all incoming live edges. We say that a protocol avoids interference if only one incoming edge is live for all receiving nodes.

In wireless networks, the relay nodes operate in either half or full-duplex mode. In case of half-duplex operation, a node cannot simultaneously listen and transmit, i.e., an incoming edge and an outgoing edge of a node cannot simultaneously be active.

In this paper, we use uppercase letters to denote matrices and lowercase letters to denote vectors/scalars. Vectors and scalars are differentiated only through the context. Irrespective of whether a particular random entity is a scalar, vector or a matrix, the entity will be represented using boldface letters.

Between any two adjacent nodes v_x, v_y of a wireless network, we assume the following channel model.

$$\mathbf{y} \ = \mathbf{H}\mathbf{x} + \mathbf{w} \ , \tag{1}$$

where $\mathbf{y}$ corresponds to the received signal at node v_y, $\mathbf{w}$ is the noise vector, $\mathbf{H}$ is a matrix and $\mathbf{x}$ is the vector transmitted by the node v_x.

2) Assumptions: We follow the literature in making the assumptions listed below. Our description is in terms of the equivalent complex-baseband, discrete-time channel.

1) All channels are assumed to be quasi-static and to experience Rayleigh fading and hence all fade coefficients are i.i.d., circularly-symmetric complex gaussian $\mathbb{C}\mathcal{N}(0,1)$ random variables.
2) The additive noise at each receiver is also modelled as possessing an i.i.d., circularly-symmetric complex gaussian $\mathbb{C}\mathcal{N}(0,1)$ distribution.
3) Each receiver (but none of the transmitters) is assumed to have perfect channel state information of all the upstream channels in the network. [2]

C. Results

In this paper, we characterize maximum diversity, maximum multiplexing gain and achievable DMT for arbitrary cooperative networks. Some of these results were presented in conference versions of this paper [21]–[24] (see also [25], [26]). Special classes of networks are considered in the second part of this two-part paper, [1]. Optimal code design for all proposed protocols in both parts of the paper can also be found there.

The principal results established in this paper are the following (see Table I-C for a tabular of results).

1) The maximum diversity of a multi-antenna multi-terminal network is equal to the value of the min-cut between the source and the destination.
2) The maximum multiplexing gain for a ss-ss full-duplex multi-antenna network is equal to the minimum rank of any cut between the source and the destination.
3) A DMT which is linear between the maximum diversity and maximum multiplexing gain is achievable for full-duplex single-antenna relay networks.

We also prove the following general results, that are useful in computing the DMT of cooperative networks

4) The colored noise encountered in cooperative networks can be treated as white for DMT computations.
5) We provide a lower bound on the DMT of triangular matrices.
6) We compute the DMT of a parallel MIMO channel in terms of the DMT of the component MIMO links.

D. Relation to Existing Literature

1) *Proof of a Conjecture by Rao and Hassibi:*
The results in *Example 5* in Section II-E.1 prove Conjecture 1 of [28] and [29].
2) *Lower bound on the DMT of various AF Protocols:* Certain results in this paper can be used to recover existing results on the DMT of AF protocols in a simpler, concise and more intuitive manner.
NAF Protocol: We compute a lower bound on the DMT of the NAF protocol, which turns out to be tight, as proved in [6].
SAF Protocol: We compute a lower bound on the DMT of the Slotted Amplify-and-Forward (SAF) protocol

[2]However, for the protocols proposed in this paper, the CSIR is utilized only at the sink, since all the relay nodes are required to simply amplify and forward the received signal.

TABLE I
PRINCIPAL RESULTS SUMMARY

Network	No of sources/ sinks	No of antennas in nodes	FD/ HD	Direct Link	Upper bound on Diversity/DMT $d_{\text{bound}}(r)$	Achievable Diversity/DMT $d_{\text{achieved}}(r)$	Is upper bound achieved?	Reference
Arbitrary	Multiple	Multiple	FD/HD	✓	$d(0) =$ Min-cut	$d(0) =$ Min-cut	✓ ($d_{\max}$ achieved)	Theorem 3.1
Arbitrary	Multiple	Multiple	FD/HD	$\times$	$d(0) =$ Min-cut	$d(0) =$ Min-cut	✓ ($d_{\max}$ achieved)	Theorem 3.1
Arbitrary	Single	Multiple	FD	✓	$r_{\max}$ = Rank of Min-cut	$r_{\max}$ = Rank of Min-cut	✓ (r_{max} achieved)	Theorem 3.4
Arbitrary Directed Acyclic Networks	Single	Single	FD	✓	Concave in general	$d_{\max}(1-r)^+$	A linear DMT between $d_{\max}$ and $r_{\max}$ is achieved	Theorem 4.1

under the relay-isolation assumption [19] in *Example 2* of Section II-E.1. From the results in [19], this lower bound is in fact tight.

N-Relay MIMO NAF Channel Appearing in [17]:
In *Example 5* of Section II-E.1, we prove an improved lower bound on the DMT for the MIMO NAF protocol for a two-hop multi-antenna network with a direct link compared to the bound in [17].

3) *The diversity of arbitrary cooperative networks.*
As noted earlier, we characterize completely the maximum diversity order attainable for arbitrary cooperative networks and it is shown that an amplify-and-forward scheme is sufficient to achieve this. Special cases of these were derived for the MIMO two-hop relay channel in [17], under a certain condition on the number of antennas (See *Corollary 1* in that paper). Also, the diversity order of layered networks using amplify-and-forward networks is characterized in [18]. The same result is obtained using lower-complexity codes in [41] and [42]. For arbitrary ss-ss networks, upper bounds on the diversity order of ss-ss networks are derived in [53], however, no achievability results are given there. Very recently, [30] have characterized the diversity of general ss-ss networks. It must be noted that this result can be obtained as a special case of our result for multi-terminal networks, which appeared in [21], although the achievability strategy is different in [30].

4) *DMT of single-antenna full-duplex networks* As a consequence of the compound channel results in [33], the optimal DMT of full-duplex single-antenna networks can be proved to be equal to the cut-set bound. While most of the results in the current paper focus on either multi-antenna or half-duplex networks, it must be noted that the schemes presented in [33] involve long random coding arguments in contrast to the short block-length, explicit schemes presented in the present paper.

5) *Maximum multiplexing gain of cooperative relay networks* The maximum multiplexing gain for single-antenna full-duplex relay networks can be readily obtained from the results in [33] and it is potentially possible to extend these results to the multiple antenna case. We adopt however, a different approach here, and utilize a conversion from the deterministic wireless network to the fading network in order to determine the MMG. The conversion is operational in the sense that a capacity achieving strategy on deterministic network can be converted into a MMG-achieving strategy for the fading network.

6) The DMT of the parallel channel in closed form is obtained in Lemma 2.5. A special case of this result is derived in [18] where the authors characterize the parallel channel DMT for the case when all the individual channels have the same DMT.

E. Outline

In Section II, we present basic results and techniques which will be of use in studying the DMT of multi-hop networks. In this section, we introduce the information-flow diagram (i-

f diagram), and prove a lower bound on the DMT of lower triangular matrices. In Section III, we characterize the extreme points of the optimal DMT of arbitrary ss-ss networks. We provide a lower bound to the DMT of arbitrary ss-ss networks with single-antenna, full-duplex relays in Section IV.

In the sequel to the present paper, we will make use of the basic results and techniques introduced here, to characterize the optimal DMT of certain classes of networks. The second part will also provide code designs for all the protocols proposed in both parts of the paper.

II. Basic Results for Cooperative Networks

We begin by reviewing the notion of DMT in point-to-point channels and then go on to explain how the DMT becomes a meaningful tool in the study of cooperative wireless networks. Later in this section, we develop general techniques, which will prove useful in deriving results on the optimal DMT of ss-ss networks.

A. Background

1) Diversity-Multiplexing Gain Tradeoff: Let R denote the rate of communication across the network in bits per network use. Let $\wp$ denote the protocol used across the network, not necessarily an AF protocol. Let r denote the multiplexing gain associated to rate R defined by

$$R \quad = \quad r\log(\rho).$$

The probability of outage for the network operating under protocol $\wp$, i.e., the probability of outage of the induced channel in (2) is then given by

$$P_{\text{out}}(\wp,R) = \inf_{\Sigma_x \,\geq\, 0,\ \text{Tr}\ (\Sigma_x)\,\leq\, n\rho} \Pr(I(\mathbf{x};\mathbf{y}) \ <\ nR|\mathbf{H}(\wp)=H(\wp)),$$

where $\mathbf{H}(\wp)$ denotes the collection of all random variables associated with the induced channel of the protocol $\wp$. Let the outage exponent $d_{\text{out}}(\wp,r)$ be defined by

$$d_{\text{out}}(\wp,r) \quad = \quad -\lim_{\rho\to\infty}\frac{P_{\text{out}}(\wp,R)}{\log(\rho)},$$

and we will indicate this by writing

$$\rho^{-d_{\text{out}}(\wp,r)} \quad \doteq \quad P_{\text{out}}(\wp,R).$$

The symbols $\dot{\geq}$, $\dot{\leq}$ are similarly defined.

The outage exponent $d_{\text{out}}(r)$ of the network associated to multiplexing gain r is then defined as the supremum of the outages taken over all possible protocols, i.e.,

$$d_{\text{out}}(r) \quad = \quad \sup_{\wp} d_{\text{out}}(\wp,r).$$

A distributed space-time code (more simply, a code) operating under a protocol $\wp$ is said to achieve a diversity gain $d(\wp,r)$ if

$$P_e(\wp,\rho) \doteq \rho^{-d(\wp,r)}\ ,$$

where $P_e(\rho)$ is the average error probability of the code $C(\rho)$ under maximum likelihood decoding. Using Fano's inequality, it can be shown (see [4]) that for a given protocol,

$$d(\wp,r) \quad \leq \quad d_{\text{out}}(\wp,r).$$

The DMT $d(r)$ of the network associated to a multiplexing gain r is then defined as the supremum of all achievable diversity gains across all possible protocols and codes.

We will refer to the outage exponent $d_{\text{out}}(r)$ of a protocol in this paper as the DMT $d(r)$ of the protocol, since for every protocol discussed in this paper, we shall identify a corresponding coding strategy that achieves $d(\wp,r)$ in the sequel [1] to the present paper.

Definition 1: Given a random matrix $\mathbf{H}$ of size $m\times n$, we define the *DMT of the matrix* $\mathbf{H}$ as the DMT of the associated channel $\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{w}$ where $\mathbf{x}$ and $\mathbf{y}$ are column vectors of size $(n\times 1)$ and $(m\times 1)$ respectively, and where $\mathbf{w}$ is a $\mathcal{CN}(0,I)$ column vector. We denote the DMT of the matrix $\mathbf{H}$ by $d_H(.)$

2) Cut-Set bound on DMT: On any network, the cut-set upper-bound on mutual information of a general multi-terminal network [49] translates into an upper bound on the DMT. This was formalized in [9] as follows:

Lemma 2.1: Let $r\log(\rho)$ be the rate of communication between the source and the sink. Given a cut ω between source and destination, let $\mathbf{H}_\omega$ denote the transfer matrix between nodes on the source-side of the cut and those on the sink-side, and let $d_\omega(r)$ be the DMT of $\mathbf{H}_\omega$. Then the DMT of communication between source and destination is upper bounded by

$$d(r)\leq\min_{\omega\in\Lambda}\{d_\omega(r)\},$$

where Λ is the set of all cuts between the source and the destination.

An example of the dominating min-cut is shown in Fig. 2.

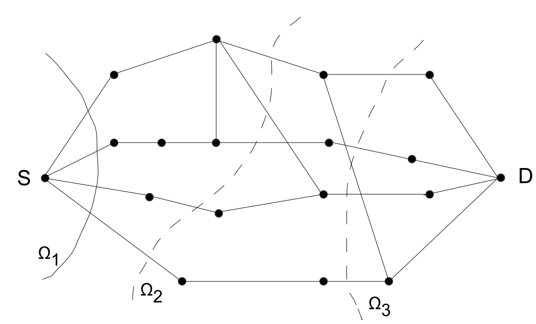


Fig. 2. Cuts in a network. Here, the min-cut is Ω_1.

3) Amplify and Forward Protocols: By an AF protocol $\wp$, we will mean a protocol $\wp$ in which each node in the network operates in an amplify-and-forward fashion. Such protocols induce a linear channel model between source and sink of the form:

$$\mathbf{y} \ = \mathbf{H}(\wp)\mathbf{x}+\mathbf{w}\ , \tag{2}$$

where $\mathbf{y}\in\mathbb{C}^m$ denotes the signal received at the sink, $\mathbf{w}$ is the noise vector, $\mathbf{H}(\wp)$ is the $(m\times n)$ induced channel matrix and $\mathbf{x}\in\mathbb{C}^n$ is the vector transmitted by the source. We impose the following energy constraint on the vector $\mathbf{x}$ transmitted by

the source,

$$\text{Tr }(\Sigma_x) \ := \ \text{Tr }(\mathbb{E}\{\mathbf{x}\mathbf{x}^\dagger\}) \quad \leq \quad n\rho,$$

where Tr denotes the trace operator. We will assume a symmetric energy constraint at the relays as well as the source. Assuming the noise power spectral density to be equal to 1, ρ corresponds to the SNR for any individual link. We consider both half and full-duplex operation at the relay nodes.

Our attention here will be restricted to amplify-and-forward (AF) protocols since as we shall see, this class of protocols can often achieve the DMT of a network. More specifically, our protocol will require the links in the network to operate according to a schedule which determines the time slots during which a node listens as well as the time slots during which it transmits. When we say that a node listens, we will mean that the node stores the corresponding received signal in its buffer. When a node does transmit, the transmitted signal is simply a scaled version of the most recent received signal contained in its buffer, with the scaling constant chosen to meet a transmit power constraint. [3] In particular, nodes in the network are not required to decode and then re-encode. It turns out [6] that the value of the scaling constant does not affect the DMT of the network operating under the specific AF protocol. Without loss of accuracy therefore, we will assume that this constant is equal to 1. It follows that, for any given network, we only need specify the schedule to completely specify the protocol. This will create a virtual MIMO channel of the form $\mathbf{y} \ = \ \mathbf{H}\mathbf{x} + \mathbf{w}$ where $\mathbf{H}$ is the effective transfer matrix and $\mathbf{w}$ is the noise vector, which is in general colored.

In following subsections of this section, we will develop techniques to handle colored noise as well as establish results on the DMT of some elementary network connections. We will also establish lower bounds on the DMT of lower triangular matrices, which will be useful later in computing the DMT of certain protocols. We will also establish the maximum multiplexing gain for channel matrices possessing certain structure.

B. White in the Scale of Interest

In this section, we provide two results that will be extensively used in all future sections: Theorem 2.3, which states that noise, even though correlated, can be treated as white in the scale of interest and Lemma 2.4, which proves that i.i.d. gaussian inputs are sufficient to attain the outage exponent of any channel of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$.

If $\mathbf{h}$ is a Rayleigh random variable, then it is very easy to see that, for any given ϵ and ρ,

$$Pr\{|\mathbf{h}|^2 > \rho^\epsilon\} \leq \exp(-\rho^\epsilon).$$

Interestingly, a similar statement holds even when we replace h by a polynomial in several Rayleigh random variables.

Lemma 2.2: Let $\{\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M\}$ be a collection of i.i.d. Rayleigh random variables. Let $f \in \mathbb{C}[X_1, X_2, ..., X_M]$ be a polynomial in the variables X_i without a constant term. Then there exists $A > 0, B > 0, d > 0, \delta > 0$ such that

$$Pr\{|f(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > k\} \quad \leq \quad A\exp(-Bk^{\frac{1}{d}}), \forall\ k \geq \delta,$$

where the constants A, B, d, δ are independent of k.

Proof: See Appendix I ■

We are now ready to establish that if the noise covariance matrix has a certain structure, then it can be considered as white noise for the purpose of DMT computation.

Theorem 2.3: Consider a channel of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$. Let $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_L$ be L i.i.d., Rayleigh random variables. Let $\mathbf{G}_i, i = 1, 2, .., M$ be $N \times N$ matrices in which each entry is a polynomial function of the random variables $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_L$. Let $\mathbf{z} = \mathbf{z}_0 + \sum_{i=1}^{M} \mathbf{G}_i \mathbf{z}_i$ be the noise vector for a channel of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. Let $\{\mathbf{z}_i\}$ be independent $\mathbb{C}\mathcal{N}(\underline{0}, I)$ random vectors. Let the random matrix $\mathbf{H}$ be a function of the random variables $\mathbf{h_i}$. Then the noise vector $\mathbf{z}$ is white in the scale of interest, i.e., the DMT of the channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ is the same as the DMT of the channel $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ with $\mathbf{w}$ being a $\mathbb{C}\mathcal{N}(\underline{0}, I)$ random vector.

Proof: See Appendix II. ■

Lemma 2.4: [4] For any channel that is of the form $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ with $\mathbf{w}$ being white gaussian noise, i.i.d. gaussian inputs are sufficient to attain the best possible outage exponent of the channel.

Proof: While a complete proof is available in [4], we provide a sketch of the same proof for the sake of completeness. The outage probability is given by,

$$P_{\text{out}}(R) \quad = \quad \inf_{\Sigma_x:\ \text{Tr }(\Sigma_x) \leq \mathbb{P}} \Pr\{I(\mathbf{x}; \mathbf{y} \mid \mathbf{H} = H) < R\}$$

The mutual information is a function of the channel realization and the distribution of the input. Nevertheless, without loss of optimality, the distribution can be chosen to be gaussian, leading to

$$P_{\text{out}}(R) \quad = \quad \inf_{\Sigma_x:\ \text{Tr }(\Sigma_x) \leq \mathbb{P}} \Pr\{\log\det(I + \rho\mathbf{H}\Sigma_x\mathbf{H}^\dagger) < R\}.$$

By bounding the eigenvalues of Σ_x, the outage probability can be bounded below and above as,

$$\begin{aligned} &\Pr\{\log\det(I + \frac{\rho}{m}\mathbf{H}\mathbf{H}^\dagger) \leq R\} \\ &\geq \quad P_{\text{out}}(R) \geq \Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^\dagger) \leq R\}. \end{aligned}$$

As $\rho \to \infty$, it can be shown that the two bounds converge so that we get Equation (9) in [4]),

$$P_{\text{out}}(R) \quad \doteq \quad P(\log\det(I + \rho\mathbf{H}\mathbf{H}^\dagger) < R). \tag{3}$$

■

The noise that we deal with in this paper will always satisfy the conditions in Theorem 2.3. Hence we will make the two assumptions appearing below throughout the paper:

- the transmitted signal has an i.i.d. gaussian distribution
- the noise is white in the scale of interest.

C. DMT of Elementary Network Connections

[3] More sophisticated linear processing techniques would include matrix transformations of the incoming signal, but turn out to be not needed here.

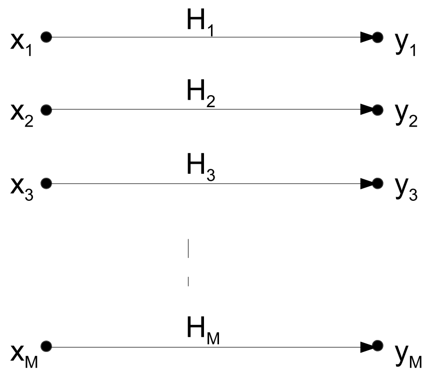


Fig. 3. The Parallel channel with M sub-channels

1) Parallel Network : The lemma below presents an expression for the DMT of a parallel channel in terms of the DMT of the individual links.

Lemma 2.5: Consider a parallel channel with M links, with the ith link having representation $\mathbf{y_i} = \mathbf{H_i x_i} + \mathbf{w_i}$, and let $d_i(\cdot)$ denote the corresponding DMT. Then the DMT of the overall parallel channel is given by

$$d(r) = \inf_{(r_1, r_2, \cdots, r_M):\ \sum_{i=1}^{M} r_i = r} \sum_{i=1}^{M} d_i(r_i). \tag{4}$$

Proof: See Appendix V. ■

The following lower and upper bounds on the outage exponent are immediate from (4):

$$d(r) \leq \sum_{i=1}^{M} d_i\left(\frac{r}{K}\right) \tag{5}$$

$$d(r) \geq \sum_{i=1}^{M} d_i(r) \tag{6}$$

To determine the DMT of the parallel channel when all component channels are identically distributed with a DMT that is a convex function of the rate, we will make use of the following Lemma from the theory of majorization [48]:

Lemma 2.6: [48] If $f(.)$ is a symmetric function in variables $r_1, r_2, \ldots, r_N$ and is convex in each of the variables $r_i, i = 1, 2, \ldots, N$, then,

$$\inf_{(r_1, r_2, \cdots, r_N):\ \sum_{i=1}^{N} r_i = r} f(r_1, r_2, ..., r_N) = f\left(\frac{r}{N}, \frac{r}{N}, ..., \frac{r}{N}\right) \tag{7}$$

The corollary below follows as a result.

Corollary 2.7: The DMT of a parallel channel with all the individual channels being identical and having a convex DMT is given by:

$$d(r) = M d_1\left(\frac{r}{M}\right). \tag{8}$$

The result in Corollary 2.7 was also obtained in [17].

2) Parallel Channel with Repeated Coefficients:

Lemma 2.8: Consider a parallel channel with M links and repeated channel matrices. More precisely, let there be N distinct channel matrices $H^{(1)}, H^{(2)}, ..., H^{(N)}$, with $H^{(i)}$ repeating in n_i sub-channels, such that $\sum_{i=1}^{N} n_i = M$.

Then the DMT of such a parallel channel is given by,

$$d(r) = \inf_{(r_1, r_2, \cdots, r_M):\ \sum_{i=1}^{N} n_i r_i = r} \sum_{i=1}^{N} d_i(r_i). \tag{9}$$

Proof: The proof is along the lines of the proof of Lemma 2.5, and is given in Appendix VI. ■

D. Maximum Multiplexing gain

In this section, we derive the maximum multiplexing gain (MMG) of a MIMO channel matrix with each entry of the matrix being a polynomial function of certain Rayleigh random variables. We begin by deriving certain properties of polynomial functions of gaussian random variables and we will later use these characteristics to obtain the MMG.

Lemma 2.9: Let $p \in \mathbb{R}[X]$ be any non-constant polynomial, and let its degree be d. Consider the set $\mathcal{R}$ of all $x \in \mathbb{R}$ over which the following two conditions are satisfied:

$$|p(x)| \leq k, \tag{10}$$

$$|p'(x)| \geq m. \tag{11}$$

This subset $\mathcal{R}$ of $\mathbb{R}$ can be expressed as the union

$$\mathcal{R} = \cup_{i=1}^{L} R_i \tag{12}$$

of disjoint intervals $R_i = [a_i, b_i]$. Furthermore, $L \leq 2d$.

Proof: See Appendix III ■

Lemma 2.10: Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$ be a collection of independent gaussian random variables. Let $f \in \mathbb{R}[X_1, X_2, ..., X_N]$ be a polynomial in the variables X_i. Then there exists constants $A > 0, d > 0, K > 0$ such that

$$\Pr\{|f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)| < \delta\} \leq A\delta^{\frac{1}{d}}, \ \ \forall\ 0 \leq \delta < K,$$

where the constants A, d, K depend only on f and not on δ.

Proof: See Appendix IV ■

We will proceed to utilize this lemma to obtain the MMG.

Definition 2: Given a random matrix $\mathbf{H}$, which is a function of random variables $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N$, we define the structural rank of $\mathbf{H}$ as the maximum rank attained by $\mathbf{H}$, where the maximum is computed over all possible realizations of the $\{\mathbf{h}_i\}$. We denote the structural rank of a random matrix $\mathbf{H}$ by $\mathbb{R}\text{ank}(\mathbf{H})$.

Theorem 2.11: Consider a channel of the form $\mathbf{y} = \mathbf{Hx} + \mathbf{w}$, where $\mathbf{H} \in \mathbb{C}^{N \times N}$ is a random matrix, and $\mathbf{x}, \mathbf{y}, \mathbf{w}$ are N-length column vectors representing the transmitted signal, received signal and the noise vector respectively, with the noise being white in the scale of interest. If the entries of $\mathbf{H}$ are polynomial functions of certain underlying Rayleigh random variables, then the maximum multiplexing gain D of the channel is given by,

$$D = \mathbb{R}\text{ank}(\mathbf{H}).$$

Proof: We will prove that the MMG of the channel is equal to the structural rank $\mathbb{R}\text{ank}(\mathbf{H}) =: m$ of $\mathbf{H}$. Clearly, for any given $\mathbf{H}$, the MMG is upper-bounded by the rank of $\mathbf{H}$, which is lesser than m. Therefore the upper-bound of m on the MMG is clear. Next, we will show that a MMG of m is achievable, i.e., for any $\delta > 0$, a multiplexing gain of $(m-\delta)$ yields a non-zero diversity gain.

Consider transmission at a multiplexing gain of $r = m - \delta$. Since $\mathbf{H}$ is of structural rank m, there is a $m \times m$ sub-matrix H_m of structural rank m. Then $\mathbf{H}_m\mathbf{H}_m^\dagger$ is a principal sub-matrix of $\mathbf{H}\mathbf{H}^\dagger$. Using the inclusion principle (*Theorem* 4.3.15 in [50]) and the fact that only m eigenvalues of $\mathbf{H}$ are non-zero, we obtain that,

$$\log\det(I + \rho\mathbf{H}\mathbf{H}^\dagger) \geq \log\det(I + \rho\mathbf{H}_m\mathbf{H}_m^\dagger).$$

Therefore, we get the outage exponent for rate $r = m - \delta$ as

$$\begin{aligned}
\rho^{-d_{out}(r)} &= \Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^\dagger) < (r)\log\rho\} \\
&\leq \Pr\{\log\det(I + \rho\mathbf{H}_m\mathbf{H}_m^\dagger) < (r)\log\rho\} \\
\rho^{-d_{out}(m-\delta)} &\leq \Pr\{\det(I + \rho\mathbf{H}_m\mathbf{H}_m^\dagger) < \rho^{(m-\delta)}\} \\
&\leq \Pr\{\det(\rho\mathbf{H}_m\mathbf{H}_m^\dagger) < \rho^{(m-\delta)}\} \\
&= \Pr\{\det(\mathbf{H}_m\mathbf{H}_m^\dagger) < \rho^{-\delta}\} \\
&= \Pr\{|\det(\mathbf{H}_m)|^2 < \rho^{-\delta}\}.
\end{aligned}$$

Let the random matrix $\mathbf{H}$, and thereby its sub-matrix $\mathbf{H}_m$, be a function of the Rayleigh random variables $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N$. Let us denote the real and imaginary parts of this collection of Rayleigh random variables by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, where $N = 2M$. Now $\mathbf{x}_i$ are i.i.d. gaussian random variables, i.e., they are distributed as $\mathcal{N}(0,1)$. Then $|\det(\mathbf{H}_m)|^2$ is a non-zero real polynomial $p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ in $\mathbf{x}_i$. Since $p = |\det(\mathbf{H}_m)|^2$ is positive, $|p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)| = p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$.

We can now use Lemma 2.10 to obtain that

$$\begin{aligned}
\Pr\{|\det(\mathbf{H}_m)|^2 < \rho^{-\delta}\} &= \Pr\{|p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)| < \rho^{-\delta}\} \\
&\leq A\rho^{-\delta/d}, \qquad (13)
\end{aligned}$$

for some positive constants A, d, K with $\rho^{-\delta} < K$. Let $\rho_0^{-\delta} = K$. Then we can see that (13) is valid for all $\rho > \rho_0$.

This leads to,

$$\begin{aligned}
\rho^{-d_{out}(m-\delta)} &\leq A\rho^{-\delta/d}, \quad \forall\rho > \rho_0 \\
\Rightarrow \rho^{-d_{out}(m-\delta)} &\dot{\leq} \rho^{-\delta/d} \\
\Rightarrow d_{out}(m-\delta) &\geq \delta/d \\
&> 0.
\end{aligned}$$

Thus a MMG of m is achievable and this concludes the proof. ∎

E. A Lower Bound on the DMT of Block-Lower-Triangular Matrices

In this section, we give a lower bound on the DMT of "block-lower-triangular"(blt) matrices, that are defined below. In many situations, the matrices induced by AF protocols in a ss-ss network will turn out to posses block-lower-triangular structure.

Definition 3: Consider a set of $N_i \times N_j$ matrices $A_{ij}, j = 1, 2, ..., N, i \geq j$. Let A be the blt matrix comprised of the block matrices A_{ij} in the (i,j)th position and zeros elsewhere, i.e.,

$$A = \begin{bmatrix} A_{11} & 0 & \ldots & 0 \\ A_{21} & A_{22} & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ A_{N1} & A_{N2} & \ldots & A_{NN} \end{bmatrix}.$$

We define the l-th sub-diagonal matrix, $A^{(\ell)}$ of such a blt matrix A as the matrix comprising only of the entries $A_{\ell 1}, A_{(\ell+1)2}, ..., A_{(\ell+N-1)N}$ with zeros everywhere else i.e.,

$$(A^{(\ell)})_{ij} = \begin{cases} A_{ij} & \text{if} \quad i - j = l - 1, \\ 0_{N_i \times N_j} & \text{otherwise.} \end{cases}$$

The *last* sub-diagonal matrix of A is defined as the sub-diagonal matrix $A^{(\ell)}$ of A, where ℓ is the largest integer for which $A^{(\ell)}$ is non-zero. Thus, for example, the matrix whose only nonzero terms are the diagonal entries of A corresponds to $A^{(\ell)}$ with $\ell = 0$ and the matrix whose only nonzero entry is A_{N1} corresponds to $A^{(\ell)}$ with $\ell = (N-1)$.

The theorem below establishes lower bounds on the DMT of channel matrices which have a blt structure.

Theorem 2.12: Consider a random blt matrix $\mathbf{H}$ having component matrices $\mathbf{H}_{ij}$ of size $N_i \times N_j$. Let $M := \sum_{i=1}^{N} N_i$ be the size of the square matrix $\mathbf{H}$.

Let $\mathbf{H}^{(0)}$ be the diagonal part of the matrix $\mathbf{H}$ and $\mathbf{H}^{(\ell)}$ denote the last sub-diagonal matrix of $\mathbf{H}$, as given by Definition 3. Then,

1) $d_H(r) \geq d_{H^{(0)}}(r)$.
2) $d_H(r) \geq d_{H^{(\ell)}}(r)$.
3) In addition, if the entries of $H^{(\ell)}$ are independent of the entries in $H^{(0)}$, then $d_H(r) \geq d_{H^{(0)}}(r) + d_{H^{(\ell)}}(r)$.

Proof: The channel is given by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. Since the noise is white in the scale of interest, by Theorem 2.3, the DMT of this channel is the same as that of a channel with the noise distributed as $\mathbb{CN}(0, I)$. Therefore, without loss of generality, we assume that $\mathbf{w}$ is distributed as $\mathbb{CN}(0, I)$.

For any given matrix H, the outage probability exponent [4] is given by

$$\rho^{-d_H(r)} \doteq \inf_{\Sigma_x : \text{Tr } (\Sigma_x) \leq \mathbb{P}} \Pr\{I(\mathbf{x}; \mathbf{y} : \mathbf{H} = H) \leq r\log\rho\}.$$

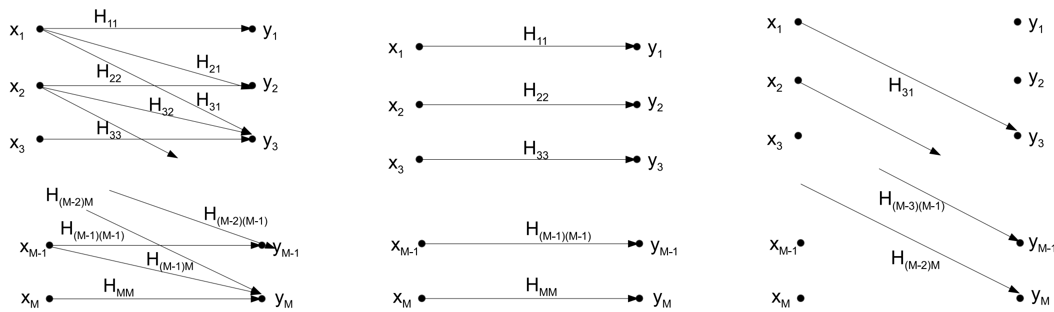
To estimate this exponent, we begin by identifying lower bounds on the mutual information. Note that by Lemma 2.4, for the purposes of computing outage exponent, we may assume without loss of optimality that the input $\mathbf{x}$ is distributed as $\mathbb{C}\mathcal{N}(0, I)$. We will make this assumption.

Due to the fact that the last sub-diagonal matrix is given by the ℓ-th sub-diagonal matrix, we have,

$$\mathbf{y}_i = \mathbf{H}_{ii}\mathbf{x}_i + \mathbf{H}_{i(i-1)}\mathbf{x}_{i-1} + ... + \mathbf{H}_{i(i-\ell)}\mathbf{x}_{i-\ell} + \mathbf{w}_i.$$

Starting with the mutual information term, we have,

$$I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H) = \sum_{i=1}^{M} I(\mathbf{x}_i\mathbf{y}|\mathbf{H} = H, \mathbf{x}_1^{i-1}). \qquad (14)$$



(a) The i-f diagram (b) Showing only the direct links (c) Showing only cross links.

Fig. 4. The i-f diagram for the block lower triangular matrix and its decomposition.

Consider next, the following series of inequalities for all $i = 1, \ldots, N$:

$$\begin{aligned} & I(\mathbf{x}_i;\mathbf{y}|\mathbf{H}=H,\mathbf{x}_1^{i-1}) \\ \geq \;& I(\mathbf{x}_i;\mathbf{y}_i|\mathbf{H}=H,\mathbf{x}_1^{i-1}) \\ = \;& I(\mathbf{x}_i;\mathbf{H}_{ii}\mathbf{x}_i+\mathbf{H}_{i(i-1)}\mathbf{x}_{i-1} \\ & +\ldots+\mathbf{H}_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{H}=H,\mathbf{x}_1^{i-1}) \\ = \;& I(\mathbf{x}_i;H_{ii}\mathbf{x}_i+H_{i(i-1)}\mathbf{x}_{i-1} \\ & +\ldots+H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{H}=H,\mathbf{x}_1^{i-1}) \\ = \;& I(\mathbf{x}_i;H_{ii}\mathbf{x}_i+H_{i(i-1)}\mathbf{x}_{i-1} \\ & +\ldots+H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{x}_1^{i-1}) \\ = \;& I(\mathbf{x}_i;H_{ii}\mathbf{x}_i+\mathbf{w}_i|\mathbf{x}_1^{i-1}) \\ = \;& I(\mathbf{x}_i;H_{ii}\mathbf{x}_i+\mathbf{w}_i). \end{aligned}$$

The last step follows since $\{\mathbf{x_i}\}$ are independent under the assumed $\mathbb{CN}(0,I)$ distribution. We thus have,

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\mathbf{H}=H) \;&\geq\; \sum_{i=1}^{M} I(\mathbf{x}_i;H_{ii}\mathbf{x}_i+\mathbf{w}_i) \\ &\geq\; I(\mathbf{x};\mathbf{H^{(0)}}\mathbf{x}+\mathbf{w}|\mathbf{H^{(0)}}=H^{(0)}). \end{aligned} \tag{15}$$

In the above, as is customary, whenever a variable with a negative index is encountered, it should be interpreted as if the variable were not present. From (15), it follows that

$$\begin{aligned} \rho^{-d_H(r)} \;&\doteq\; \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}=H)\leq r\log\rho\} \qquad (16) \\ &\leq\; \Pr\{I(\mathbf{x};\mathbf{H}^{(0)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(0)}=H^{(0)})\leq r\log\rho\} \\ &\doteq\; \rho^{-d_{H^{(0)}}(r)}, \end{aligned}$$

i.e.,

$$d_H(r) \;\geq\; d_{H^{(0)}}(r). \tag{17}$$

In the "information-flow diagrams" appearing in Fig. 4, the lower bounding of the mutual information by replacing the matrix H by the diagonal $H^{(0)}$ can be seen to correspond to a pruning of the graph shown in Fig. 4(a) resulting in the figure in Fig. 4(b).

Similarly, we have a second set of inequalities which correspond effectively to replacing the matrix H by the last sub-diagonal matrix $H^{(\ell)}$. This corresponds to the pruned graph appearing in Fig. 4(c).

$$\begin{aligned} & I(\mathbf{x}_{i-\ell};\mathbf{y}|\mathbf{H}=H,\mathbf{x}_{i-\ell+1}^N) \\ \geq \;& I(\mathbf{x}_{i-\ell};\mathbf{y}_i|\mathbf{H}=H,\mathbf{x}_{i-\ell+1}^N) \\ = \;& I(\mathbf{x}_{i-\ell};\mathbf{H}_{ii}\mathbf{x}_i+\mathbf{H}_{i(i-1)}\mathbf{x}_{i-1}+\ldots+ \\ & \mathbf{H}_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{H}=H,\mathbf{x}_{i-\ell+1}^N) \\ = \;& I(\mathbf{x}_{i-\ell};H_{ii}\mathbf{x}_i+H_{i(i-1)}\mathbf{x}_{i-1}+\ldots+ \\ & H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{H}=H,\mathbf{x}_{i-\ell+1}^N) \\ = \;& I(\mathbf{x}_{i-\ell};H_{ii}\mathbf{x}_i+H_{i(i-1)}\mathbf{x}_{i-1}+\ldots+ \\ & H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{x}_{i-\ell+1}^N) \\ = \;& I(\mathbf{x}_{i-\ell};H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i|\mathbf{x}_{i-\ell+1}^N) \\ = \;& I(\mathbf{x}_{i-\ell};H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i) \end{aligned}$$

$$\begin{aligned} \Rightarrow\; & I(\mathbf{x};\mathbf{y}|\mathbf{H}=H) \\ &= \sum_{i=N}^{1} I(\mathbf{x}_i;\mathbf{y}|\mathbf{H}=H,\mathbf{x}_{i+1}^N) \\ &\geq \sum_{i=N}^{l+1} I(\mathbf{x}_{i-\ell};\mathbf{y}|\mathbf{H}=H,\mathbf{x}_{i-\ell+1}^N) \\ &\geq \sum_{i=N}^{l+1} I(\mathbf{x}_{i-\ell};H_{i(i-\ell)}\mathbf{x}_{i-\ell}+\mathbf{w}_i) \\ &= I(\mathbf{x};\mathbf{H}^{(\ell)}x+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)}). \end{aligned} \tag{18}$$

Thus

$$\begin{aligned} \rho^{-d_H(r)} \;&\doteq\; \Pr\{I(\mathbf{x};\mathbf{y}|\mathbf{H}=H)\leq r\log\rho\} \qquad (19) \\ &\leq\; \Pr\{I(\mathbf{x};\mathbf{H}^{(\ell)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)})\leq r\log\rho\} \\ &\doteq\; \rho^{-d_{H^{(\ell)}}(r)} \\ \Rightarrow\; d_H(r) \;&\geq\; d_{H^{(\ell)}}(r). \qquad (20) \end{aligned}$$

It follows therefore, from (15) and (18) that

$$\begin{aligned} & I(\mathbf{x};\mathbf{y}|\mathbf{H}=H) \\ \geq \;& \max(I(\mathbf{x};\mathbf{H_d}\mathbf{x}+\mathbf{w}|\mathbf{H_d}=H_d), \\ & I(\mathbf{x};\mathbf{H}^{(\ell)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)})). \end{aligned} \tag{21}$$

This leads to

$$\begin{aligned}
\rho^{-d_H(r)} &\doteq \Pr\{I(\mathbf{x};\mathbf{y}:\mathbf{H}=H)\le r\log\rho\}\\
&\le \Pr\{\max\{I(\mathbf{x};\mathbf{H}^{(0)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(0)}=H^{(0)}),\\
&\quad I(\mathbf{x};\mathbf{H}^{(\ell)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)})\}\le r\log\rho\}\\
&= \Pr\{I(\mathbf{x};\mathbf{H}^{(0)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(0)}=H^{(0)})\le r\log\rho,\\
&\quad I(\mathbf{x};\mathbf{H}^{(\ell)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)}))\le r\log\rho\}\\
&= \Pr\{I(\mathbf{x};\mathbf{H}^{(0)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(0)}=H^{(0)})\le r\log\rho\}\\
&\quad .\ \Pr\{I(\mathbf{x};\mathbf{H}^{(\ell)}\mathbf{x}+\mathbf{w}|\mathbf{H}^{(\ell)}=H^{(\ell)})\le r\log\rho\}\\
&\doteq \rho^{-d_{H^{(0)}}(r)}\rho^{-d_{H^{(\ell)}}(r)}\\
&\doteq \rho^{-d_{H^{(0)}}(r)+d_{H^{(\ell)}}(r)}\\
\Rightarrow\ d_H(r) &\ge d_{H^{(0)}}(r)+d_{H^{(\ell)}}(r). \qquad (22)
\end{aligned}$$

where the first step comes about because of the independence of the entries in $\mathbf{H^{(0)}}$ and $\mathbf{H}^{(\ell)}$, which is indeed the case. ■

Remark 1: The following two matrix inequalities can be deduced from the proof of Theorem 2.12, with $\mathbf{H}^{(0)}$ and $\mathbf{H}^{(\ell)}$ defined as in the theorem:

$$\det(I+\rho\mathbf{H}\mathbf{H}^\dagger) \ \ge\ \det(I+\rho\mathbf{H}^{(0)}\mathbf{H}^{(0)\dagger}) \qquad (23)$$
$$\det(I+\rho\mathbf{H}\mathbf{H}^\dagger) \ \ge\ \det(I+\rho\mathbf{H}^{(\ell)}\mathbf{H}^{(\ell)\dagger}). \qquad (24)$$

Remark 2: Although the result is derived for lower triangular matrices, it also applies in a slightly more general setting. Consider a band matrix of the form given below,

$$\mathbf{H} = \begin{bmatrix} * & & * & & & \\ & * & & * & & \\ * & & * & & * & \\ & * & & * & & * \\ & & * & & * & \\ & & & * & & * \end{bmatrix},$$

where there are bands of non-zero entries, denoted by sequence of $*$. Let $\mathbf{H}_{ub}$ and $\mathbf{H}_{lb}$ denote matrices derived from $\mathbf{H}$, constituting of only the uppermost band and lowermost band respectively. They will be, respectively of the form,

$$\mathbf{H}_{ub} = \begin{bmatrix} & & * & & & \\ & & & * & & \\ & & & & * & \\ & & & & & * \\ & & & & & \\ & & & & & \end{bmatrix},\ \mathbf{H}_{lb} = \begin{bmatrix} & & & & & \\ & & & & & \\ * & & & & & \\ & * & & & & \\ & & * & & & \\ & & & * & & \end{bmatrix}.$$

Without affecting the DMT, the matrix $\mathbf{H}$ can be transformed to a blt matrix of larger size, by adding an appropriate number of all-zero rows at the top and all-zero columns to the right. Then the uppermost band of the $\mathbf{H}$ belongs the diagonal, and the lowermost band belongs to the last sub-diagonal of the new matrix. By invoking Theorem 2.12 for the new matrix, we get

1) $d_H(r)\ge d_{H_{ub}}(r)$.
2) $d_H(r)\ge d_{H_{\ell b}}(r)$.

If the entries of $\mathbf{H}_{\ell b}$ and $\mathbf{H}_{ub}$ are independent of each other, then we further have,

$$d_H(r)\ge d_{H_{ub}}(r)+d_{H_{\ell b}}(r).$$

1) Example Applications of the DMT Lower bound: In this subsection, we derive lower bounds to the DMT of two-hop networks under the operation of various existing AF protocols. One lower bound proves a conjecture by Rao and Hassibi [28], while a second is tighter than lower bounds known earlier. In the remaining instances, although the results do not add to what is already known, the derivations presented here are surprisingly simple and provide some intuitive explanation as to how these protocols achieve the DMT.

Example 1: Single relay, NAF protocol

Consider the relay network in Fig.1(b). Let $\mathbf{g}_d$, $\mathbf{g}_1$, $\mathbf{h}_1$ denote the channel coefficients along the links from source to the sink, source to the relay and relay to the sink respectively. The induced channel under the NAF protocol is given by,

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{g}_d & 0\\ \mathbf{g}_1\mathbf{h}_1 & \mathbf{g}_d\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_1\end{bmatrix}+\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2 + h_1\mathbf{v}\end{bmatrix} \qquad (25)$$

Since two time instants are used in order to obtain the equivalent channel matrix, $\mathbf{H}$, we have a rate loss by a factor of 2, and hence $d(r)=d_H(2r)$. It can be checked that the noise is white in the scale of interest. Now it is sufficient to study the DMT of the matrix H. Let

$$\mathbf{H}^{(0)} = \begin{bmatrix}\mathbf{g}_d & 0\\ 0 & \mathbf{g}_d\end{bmatrix}, \text{ and}$$
$$\mathbf{H}^{(\ell)} = \begin{bmatrix}0 & 0\\ \mathbf{g}_1\mathbf{h}_1 & 0\end{bmatrix}.$$

The fading coefficients $\mathbf{g}_d$, $\mathbf{g}_1$, $\mathbf{h}_1$ are independent and therefore $\mathbf{H}^{(0)}$ is independent of $\mathbf{H}^{(\ell)}$. Invoking Theorem 2.12 we obtain:

$$d_H(r) \ \ge\ d_{H^{(0)}}(r)+d_{H^{(\ell)}}(r)$$

The diversity gains $d_{H^{(0)}}(r)$ and $d_{H^{(\ell)}}(r)$ are easily evaluated as,

$$\begin{aligned}
d_{H^{(0)}}(r) &= \left(1-\frac{r}{2}\right)^+,\\
d_{H^{(\ell)}}(r) &= (1-r)^+.\\
\text{This leads to } d_H(r) &\ge \left(1-\frac{r}{2}\right)^+ + (1-r)^+.
\end{aligned}$$

This leads to the following estimate of the DMT $d(r)$ of the protocol:

$$\begin{aligned}
d(r) &= d_H(2r)\\
\Rightarrow d(r) &\ge (1-r)^+ + (1-2r)^+.
\end{aligned}$$

From [6] we know that this bound is indeed tight.

Remark 3: For the case of NAF protocol used with N relays, it can be shown that Theorem 2.12 can be used to obtain a lower bound on the DMT of NAF protocol as

$$d(r) \ \ge\ (1-r)^+ + N(1-2r)^+. \qquad (26)$$

This lower bound is proved to be tight for the N-relay case

as well in [6].

Example 2: Multiple relays, SAF

Consider the network in Fig.1(a) with N relays. We employ an M-slot AF protocol termed the Slotted Amplify-and-Forward (SAF) protocol and introduced in [19]. We assume that the relays are isolated from each others' transmissions (see [19] for a description). Each symbol transmitted by the source reaches the sink through the direct link, as well as through precisely one relayed path. For this relay-isolated case, the induced channel matrix for a M-slot protocol is given by a $M \times M$ matrix $\mathbf{H}$, with $\mathbf{g}_d$, the fading coefficient of the direct link, appearing along the diagonal, and with $\mathbf{g}_1, \ldots, \mathbf{g}_N$, the product coefficients on the different relay paths, appearing in repeated cyclic fashion along the first sub-diagonal. Let $M = kN + 1$ denote the slot length, for a positive integer k.

For example, in the $M = 5$, $N = 2$, $k = 2$ case, the induced channel matrix $\mathbf{H}$ is given by,

$$\mathbf{H} = \begin{bmatrix} \mathbf{g}_d & 0 & 0 & 0 & 0 \\ \mathbf{g}_1 & \mathbf{g}_d & 0 & 0 & 0 \\ 0 & \mathbf{g}_2 & \mathbf{g}_d & 0 & 0 \\ 0 & 0 & \mathbf{g}_1 & \mathbf{g}_d & 0 \\ 0 & 0 & 0 & \mathbf{g}_2 & \mathbf{g}_d \end{bmatrix}.$$

Since the channel is used for M time slots, we have the relation $d(r) = d_H(Mr)$ between the DMT of the protocol, $d(r)$, and the DMT of the matrix $d_H(r)$. We next proceed to find a lower bound on the DMT of the matrix. As before, we use $\mathbf{H}^{(0)} = \mathbf{g}_d I$ to denote the diagonal matrix associated to $\mathbf{H}$. Similarly, let $\mathbf{H}^{(\ell)}$ denote the last sub-diagonal matrix corresponding to $\mathbf{H}$. This matrix contains $\mathbf{g}_1, \ldots, \mathbf{g}_N$ repeated k times cyclically along the first sub-diagonal. By Theorem 2.12, the DMT of $\mathbf{H}$ can be lower bounded as,

$$\begin{aligned} d_H(r) &\geq d_{H^{(0)}}(r) + d_{H^{(\ell)}}(r) \\ \Rightarrow d(r) &\geq d_{H^{(0)}}(Mr) + d_{H^{(\ell)}}(Mr). \end{aligned}$$

The DMT of the matrices $\mathbf{H}^{(\mathbf{0})}$ and $\mathbf{H}^{(\ell)}$ can be easily derived as, $d_{H^{(0)}}(r) = \left(1 - \frac{r}{M}\right)^+$ and $d_{H^{(\ell)}}(r) = N\left(1 - \frac{r}{M-1}\right)^+$ leading to:

$$d(r) \geq (1-r)^+ + N\left(1 - \frac{M}{M-1}r\right)^+.$$

The right hand side is in fact shown to be equal to the DMT of the SAF protocol in [19] under the assumption that relays are isolated.

Example 3: Multiple-Antenna, Single-Relay, NAF protocol

Consider a single-relay network with the source, the relay and sink equipped with multiple antennas given by n_s, n_r and n_d respectively. We follow [17] and assume operation under the NAF protocol introduced in [6] for the single-antenna case. The channel matrix turns out to be given by,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_d & 0 \\ \mathbf{H}_r & \mathbf{H}_d \end{bmatrix}, \tag{27}$$

where $\mathbf{H_d}$ is the $n_d \times n_s$ fading matrix between source and the sink, $\mathbf{H_r}$ is the product fading matrix of an $n_r \times n_s$ Rayleigh fading matrix between the source and the relay and an $n_d \times n_r$ Rayleigh fading matrix between relay and sink. Proceeding in the same manner as in *Example 1*, we get

$$d(r) \geq d_{H_d}(r) + d_{H_r}(2r).$$

This lower bound appears as Theorem 1 in [17].

Example 4: Multiple Antenna, Multiple relays, NAF protocol
We consider a N-relay network with each node in the network having multiple antennas. Let n_s,n_i and n_d denote the number of antennas with the source, ith relay and the destination respectively.

The NAF protocol was proposed in [6] for the case of N relays, with all nodes possessing single antennas. The protocol can be viewed as using the NAF protocol for each relay separately (the protocol comprises of two slots, with the source transmitting to the relay and destination in the first slot and the relay and the source transmitting to the destination in the second slot) and then cycling through all the relays. The same protocol was used in the case of multiple antenna relays in [17]. However it is not clear that this is the optimal thing to do if each relay has different number of antennas. In that case, we might want to use the relay with more antennas more frequently in order to get a better performance.

Therefore, in this example, we consider a generalization of the NAF protocol for multiple antenna relays, where we cycle through all the relays for unequal periods of time. Specifically, we use a NAF protocol for relay i for m_i cycles. When we say we use a NAF protocol for relay i, it means that the source will first transmit to the relay during the first time instant and then in the second time instant the source and the relay will transmit to the destination. Thus a NAF protocol operated for a single relay for one cycle will take up 2 time instants. Let $M := \sum m_i$. Then the protocol operates for $2M$ time slots and the induced channel matrix $\mathbf{H}$ of size $2M \times 2M$ between source and destination will contain the direct link fading matrix $\mathbf{H}_d$ repeated along the diagonal and the first sub-diagonal will have the product matrix $\mathbf{H}_i$ corresponding to relay i repeated for $2m_i$ times.

More precisely, the relay matrix $\mathbf{H}_i$ is the product of the Rayleigh fading matrix $\mathbf{F}_i$ between the source and the ith relay and the Rayleigh fading matrix $\mathbf{G}_i$ between the ith relay and the destination. Then the DMT $d_i(r)$ of the product matrix $\mathbf{H}_i$ can be computed using the Rayleigh product channel DMT in [18].

This induces an effective channel matrix between the source and the destination, which will be of the form:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_D & 0 \\ \mathbf{H}_R & \mathbf{H}_D \end{bmatrix}, \tag{28}$$

where $\mathbf{H}_D = I_T \otimes \mathbf{H}_d$, with I_T denoting the identity matrix of size T, $\otimes$ denoting the tensor product and $\mathbf{H}_d$ denoting the $n_s \times n_d$ fading matrix corresponding to the direct link between the source and the destination. $\mathbf{H}_R$ is now a block diagonal matrix with the matrix $\mathbf{H}_i$ appearing along the diagonal for m_i times. Now the DMT of the protocol is given by

$$d(r) = d_H(2Mr) \geq d_{H^{(0)}}(2Mr) + d_{H^{(\ell)}}(2Mr)$$

by Theorem 2.12. Since $\mathbf{H}^{(0)}$ contains the diagonal element as H_d repeated $2M$ times, $d_{H^{(0)}}(2Mr) = d_{H_d}(r)$. Also $d_{H^{(\ell)}}(2Mr) = d_{H_R}(2Mr)$.

The DMT of the matrix H_R can be computed using the DMT of parallel channel with repeated coefficients (Lemma 2.8). This gives us

$$d_{H_R}(Mr) \quad = \quad \inf_{(r_1,r_2,\cdots,r_N):\ \sum_{i=1}^{N} \ f_i r_i = r} \ \sum_{i=1}^{N} d_i(r_i),$$

where $f_i = \frac{m_i}{M}$.

Thus a lower bound on $d(r)$ can be computed as

$$d(r) \geq d_{H_d}(r) + \inf_{(r_1,r_2,\cdots,r_N):\ \sum_{i=1}^{N} \ f_i r_i = 2r} \ \sum_{i=1}^{N} d_i(r_i).$$

Now since the activation durations $\{f_i\}$ for the relays are arbitrary, we can optimize the DMT over all possible $\{f_i\}$ such that $\sum f_i = 1$.

Thus we get

$$d(r) \geq \ d_{H_d}(r) + \sup_{\substack{\{(f_1, f_2, \cdots, f_N): \\ \sum_{i=1}^{N} \ f_i = 1\}}} \ \inf_{\substack{\{(r_1, r_2, \cdots, r_N): \\ \sum_{i=1}^{N} \ f_i r_i = 2r\}}} \ \sum_{i=1}^{N} d_i(r_i). \tag{29}$$

The scheme in [17] is now a special case of this protocol where all the relays are used for a equal duration of time, i.e., $f_i = 1/N$ for all i. After substituting $\theta_i := \frac{r_i}{2Nr}$, we get

$$d(r) \geq d_{H_d}(r) + \inf_{(\theta_1,\theta_2,\cdots,\theta_K):\ \sum_{i=1}^{K} \ \theta_i = 1} \ \sum_{i=1}^{K} d_i(2N\theta_i r)$$

which is indeed the formula in *Theorem 2* of [17]. However the lower bound on DMT that we have in (29) is better than the lower bound in *Theorem 2* of [17] since we allow for arbitrary periods of activation which is a more general approach.

Example 5: Multiple-Antenna, Multiple-Relay, Generalized NAF protocol

Let us now consider a N-relay network with the source and destination having n_s and n_d antennas and the relays having a single antenna each. For this network, the generalized NAF protocol was proposed in [28], where during the first T time instants, the source transmits to the N relays. Over the next T time slots, the relays transmit a linear transformation of the vector received over the prior T time slots. This induces an effective channel matrix between the source and the destination, which will be of the form:

$$\mathbf{H} \quad = \quad \begin{bmatrix} \mathbf{H}_D & 0 \\ \mathbf{H}_R & \mathbf{H}_D \end{bmatrix}, \tag{30}$$

where $\mathbf{H}_D = I_T \otimes \mathbf{H}_d$, with I_T denoting the identity matrix of size T, $\otimes$ denoting the tensor product and $\mathbf{H}_d$ denoting the $n_s \times n_d$ fading matrix corresponding to the direct link between the source and the destination. $\mathbf{H}_R$ is a $Tn_d \times Tn_s$ matrix which depends not only on the channel fading coefficients, but also on the linear transformations employed at the relays corresponding to the relaying path, which we is the effective relaying matrix.

Now, $\mathbf{H}$ is blt and therefore, we invoke Theorem 2.12 to get, $d_H(r) \geq d_{H^{(0)}}(r) + d_{H^{(\ell)}}(r)$. Now the matrix $\mathbf{H}^{(0)}$ corresponds to a block-diagonal matrix with $\mathbf{H}_D$ repeated twice along the diagonal or effectively, $\mathbf{H}_d$ repeated $2T$ times along the diagonal and clearly $\mathbf{H}^{(\ell)} = \mathbf{H}_R$. Therefore $d_{H^{(0)}}(r) = d_{H_d}(\frac{r}{2T})$.

The protocol utilizes $2T$ time instants to induce the effective channel matrix $\mathbf{H}$ and therefore the DMT of the protocol $d(r)$ can be given in terms of the DMT of the matrix $\mathbf{H}$ as $d(r) = d_H(2Tr)$. Thus,

$$\begin{aligned} d(r) \quad &= \quad d_H(2Tr) \\ &\geq \quad d_{H^{(0)}}(2Tr) + d_{H^{(\ell)}}(2Tr) \\ &= \quad d_{H^{(0)}}(2Tr) + d_{H_R}(2Tr) \\ &= \quad d_{H_d}(r) + d_{H_R}(2Tr). \end{aligned} \tag{31}$$

We will now present this DMT inequality in the language of [28]. In [28], the DMT of the effective relaying matrix, $\mathbf{H}_R$ is computed after compensating for only a rate loss of T time instants and let us call this as $d_R(r)$, i.e., $d_R(r) := d_{H_R}(Tr)$. Let us call the DMT of the direct link as $d_D(r) := d_{H_d}(r)$. Now (31) can be re-written as

$$d(r) \geq d_D(r) + d_R(2r), \tag{32}$$

which thus proves Conjecture 1 in [28].

III. Characterization of Extreme Points of DMT of Arbitrary Networks

In this section, we move on to considering multi-hop networks. We show that the min-cut is equal to the diversity for arbitrary multi-terminal networks with multi-antenna nodes irrespective of whether the relays operate under the half-duplex constraint or not. We also show for ss-ss full-duplex networks that the maximum multiplexing gain is equal to the min-cut rank. These two results put together characterize the two end-points of the DMT of full-duplex ss-ss networks.

A. Representation of Multi-Antenna Networks

In Section I, we described how a network is represented as a graph. The graph-representation of a network described in Section I does not differentiate between the case with single-antenna nodes and that with multiple-antenna nodes. We make this distinction in a new representation of network, described below. We will use this representation throughout this section.[4]

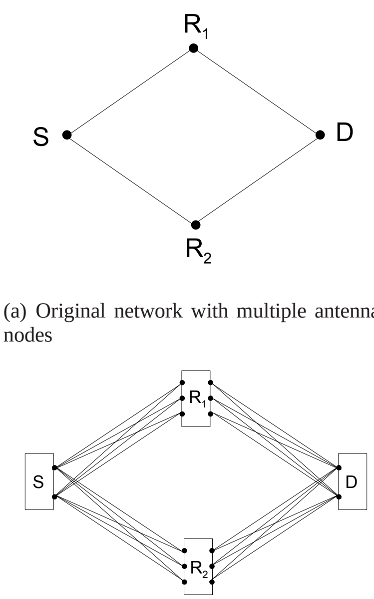
Consider a ss-ss wireless network with nodes potentially having multiple antennas. Every terminal in the network is represented by a super-node and every antenna attached to the terminal is represented by a small node associated with the super-node. There are edges drawn between small nodes of distinct super-nodes, representing communication channel between antennas of different terminals. Thus every edge

[4] A similar representation for deterministic networks is used in [31], albeit in a context different from multiple antenna nodes.

is associated with a scalar fading coefficient. Since we are dealing with wireless networks, we assume that the broadcast and interference constraints hold. In effect the vector $\mathbf{y}_i$ received by a super-node i with m_i antennas can be given in terms of the transmitted vectors $\mathbf{x}_i$ by

$$\mathbf{y_i} = \sum_{j \,\in\, \mathrm{In}(i)} H_{ij}\mathbf{x_j} + \mathbf{w_i},$$

where $\mathbf{y_i}$ and $\mathbf{w_i}$ are m_i length column column vectors, $\mathbf{x_j}$ is a m_j length vector and H_{ij} is a $m_i \times m_j$ transfer matrix between the super-node i and super-node j, containing entries with $\mathbb{C}\mathcal{N}(0,1)$ distribution. Every cut ω in the network is associated with a channel matrix, which we will denote by $\mathbf{H}_\omega$. Fig. 6 illustrates this representation for the case of a single source S, two relays R_1 and R_2 and a sink D.



(a) Original network with multiple antenna nodes

(b) Equivalent network with single antenna nodes

Fig. 5. Source and sink have 2 antennas each, relays have 3 each

It must be noted that even wireline networks can be converted into the above model of wireless networks. This can be done by adding as many number of small nodes in a super node as the number of edges emanating from or arriving at a node. Then, by making the coefficients of chosen edges to zero (or equivalently by removing corresponding edges from the representation), the broadcast and interference constraints can be nullified. Thus the class of wireline networks are naturally embedded in the class of wireless networks in the above representation.

B. Min-cut equals Diversity

Theorem 3.1: Consider a multi-terminal fading network with nodes having multiple antennas with edges connecting antennas on two different nodes having i.i.d. Rayleigh-fading coefficients. The maximum diversity achievable for any flow is equal to the min-cut between the source of the flow and the corresponding sink. Each flow can achieve its maximum diversity simultaneously.

Proof: We first consider the case where there is only a single source-sink pair. We will handle the case of single and multiple-antenna nodes separately.

Case I: Network with single antenna nodes

Let the source be S_i and sink be D_j. Let Λ_{ij} denote the set of all cuts between S_i and D_j.

From the cut-set upper bound on DMT (see Lemma 2.1),

$$\begin{aligned} d(r) &\leq \min_{\Omega\in\Lambda_{ij}} d_\Omega(r) \\ \Rightarrow d(0) &\leq \min_{\Omega\in\Lambda_{ij}} d_\Omega(0) \\ &=: \min_{\Omega\in\Lambda_{ij}} m_\omega, \end{aligned}$$

where m_ω is the number of edges crossing from the source side to the sink side in the cut ω. So, now $d(0) \leq m$, where $m := \min_\omega m_\omega$ is the min-cut.

It suffices to prove that a diversity order equal to m is achievable. We know from Menger's theorem in graph theory (see for eg. [51]), that the number of edges in the min-cut is equal to the maximum number of edge-disjoint paths between source and the sink. Schedule the network in such a way that each edge in a given edge-disjoint path is activated one by one. The same is repeated for all the edge-disjoint paths. Let the number of edges in the ith edge-disjoint path be n_i. The jth edge in the the ith edge-disjoint path is denoted by e_{ij} and the associated fading coefficient be $\mathbf{h}_{ij}$. Now define $\mathbf{h}_i := \prod_{j=1}^{n_i} \mathbf{h}_{ij}, i = 1, 2, ..., m$. So the activation schedule will be as follows: $e_{11}, e_{12}, \cdots, e_{1(n_1)}, e_{21}, \cdots, e_{2(n_2)}, \cdots, e_{m1}, e_{m2}, \cdots, e_{m(n_m)}$, where each edge is activated one at a time. The total number of time slots required for the protocol is $N := \Sigma_{i=1}^{m} n_i$. This in effect creates a parallel channel between the source S_i and destination D_j. The parallel channel contains m links, with the fading coefficients $\mathbf{h}_i$ on the link i. With this protocol in place, the equivalent channel seen by a symbol is

$$\mathbf{H} \;=\; \begin{bmatrix} \mathbf{h}_1 & 0 & \dots & & 0 \\ 0 & \mathbf{h}_2 & 0 & & 0 \\ \vdots & & \vdots & \ddots & \\ 0 & \dots & \dots & & \mathbf{h}_m \end{bmatrix}.$$

This is a parallel channel with all the channels being independent of each other and the DMT of the channels being identical. Therefore we can use Corollary 2.7 and obtain the DMT of the parallel channel as

$$d_H(r) \;=\; (m-r)^+. \tag{33}$$

This DMT can be achieved by using a DMT optimal parallel channel code.

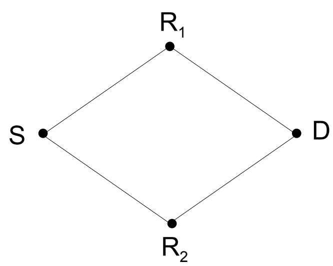
The protocol utilizes N time instants to induce this effective channel matrix, and therefore, the DMT of the protocol can be given in terms of the DMT of the channel matrix as

$$\begin{aligned} d(r) &= d_H(Nr) & (34)\\ &= (m-Nr)^+. & (35) \end{aligned}$$

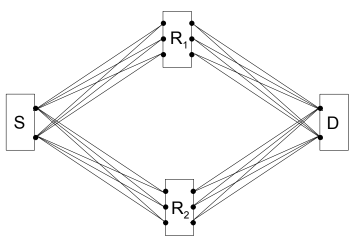
Hence the maximum achievable diversity is m.

Case II: Network with multiples antenna nodes

In the multiple antenna case, we pass on to the new representation described in Section III-A. We regard any link between a n_t transmit and n_r receive antenna as being composed of $n_t n_r$ links, with one link between each transmit and each receive antenna. Note that it is possible to selectively activate precisely one of the $n_t n_r$ Tx-antenna-Rx-antenna pairs by appropriately transmitting from just one antenna and listening at just one Rx antenna. As is to be expected, in this modified representation, a cut is defined as separating super-nodes into two sets since super-nodes represent distinct terminals. With this modification, the same strategy as in the single antenna case can then be applied to achieve a diversity equal to the min-cut in the network.



(a) Original network with multiple antenna nodes



(b) Equivalent network with single antenna nodes

Fig. 6. Illustration: $n_S = n_D = 2, n_1 = n_2 = 3$

Fig. 6 illustrates this conversion for the case of a single source S, two relays R_1 and R_2 and a sink D. Having converted the multiple antenna network into one with single antenna nodes, *Case II* follows from *Case I*. For the example shown in the figure, the min-cut and therefore the diversity is equal to 12.

Thus the proof is complete for the single flow from S_i to D_j.

When there are multiple flows in the network, we simply schedule the data of all the flows in a time-division manner. This will entail a rate loss - however, since we are interested only in the diversity, we can still achieve each flow's maximum diversity simultaneously. ■

C. Maximum Multiplexing Gain equals Minimum Rank

In this section, we determine the maximum multiplexing gain (MMG) for multi-antenna ss-ss networks to be equal to the min-cut rank (which will be formally defined later). For ss-ss networks with single-antennas, the MMG is lesser than one, because the source has a single antenna and the cut with source at one side and the rest of the nodes on the other side will yield an upper bound on MMG as one. It is possible to attain the optimal MMG of 1 by activating one path between the source to the destination either using amplify-and-forward or a decode-and-forward strategy. However, the MMG-optimal strategy becomes unclear when the number of antennas is greater than 1.

We use results from a recent work on deterministic wireless networks [32] to arrive at strategies for achieving the maximum multiplexing gain of a fading network. The achievability strategies for deterministic wireless networks are lifted to fading networks using simple algebraic techniques. We begin with discussing a new representation for ss-ss networks, potentially having multiple antenna nodes, which will be used in this section.

1) Linear Deterministic Wireless networks : In defining deterministic[5] wireless networks, we follow [31]. Every terminal in the network is represented by a super-node and each node possesses q small nodes associated with the super-node. All operations take place over a fixed finite field $\mathbb{F}_p$. There are edges drawn between small nodes of distinct super-nodes, representing communication channel between antennas of different terminals. Since we are dealing with deterministic wireless networks, we assume that the broadcast and interference constraints hold. In effect the vector $\mathbf{y}_i$ received by a super-node i can be given in terms of the transmitted vectors $\mathbf{x}_j$ of various nodes by

$$\mathbf{y_i} = \sum_{j \ \in \ \text{In}(i)} G_{ij}\mathbf{x_j},$$

where $\mathbf{y_i}$ and $\mathbf{x_j}$ are q length column column vectors in $\mathbb{F}_p$, and G_{ij} is a $m_i \times m_j$ transfer matrix between the super-node i and super-node j, taking values in $\mathbb{F}_p$. Every cut ω in the deterministic network is associated with a channel matrix, which we will denote by G_ω.

The network model of linear deterministic networks thus described has close similarities with representation of ss-ss fading networks described in Section III-A with the multiple antennas taking the place of small nodes in the case of fading networks. The difference between the two are only that deterministic network has noise-free links in comparison to the noisy links in the fading case, and that every edge coefficient is a finite field element in the deterministic network, in place of complex fading coefficient. In deterministic networks, each node transmits a q-tuple over the finite field. The theorem below from [31], computes the capacity [6] of a ss-ss linear deterministic wireless network.

Theorem 3.2: [31] Given a linear deterministic ss-ss wireless network over any finite field $\mathbb{F}_p$, $\forall\ \epsilon > 0$, the capacity C of such a relay network is given by,

$$C = \min_{\omega \ \in \ \Omega} \text{rank}(G_\omega),$$

[5]By deterministic network, we will always mean linear deterministic network.

[6]We use the term capacity to signify ϵ-error capacity, as is conventional.

where the capacity is specified in terms of the number of finite field symbols per unit time. A strategy utilizing only linear transformations over $\mathbb{F}_p$ at the relays is sufficient to achieve this capacity.

The capacity-achieving strategy in [31] utilizes matrix transformations of the input vector received over a period of T time slots at each relay. This process continues for L blocks, therefore the total number of time instants required for the scheme is $M := LT$. The achievability shows the existence of relay matrices A_i at each relay node $i \in |\mathcal{V}|$, where $\mathcal{V}$ is the set of vertices in the graph. A_i is of size $qT \times qT$, and it represents the transformation between the received vector of size qT to the vector of size qT that is transmitted.

The multi-cast version of Theorem3.2 is reproduced below:

Theorem 3.3: [31] Given a linear deterministic single-source D-sink multi-cast wireless network, $\forall\ \epsilon > 0$, the capacity C of such a network is given by,

$$C = \min_{j=1,2,..,D} \min_{\omega \in\ \Omega_j} \text{rank}(G_\omega).$$

where Ω is the set of all cuts between the source and destination j. A strategy utilizing only linear transformations at the relays is sufficient to achieve this capacity.

2) MMG of ss-ss networks: The main result of this section is given below.

Theorem 3.4: Given a ss-ss multi-antenna wireless network, with Rayleigh fading coefficients, the MMG of the network is given by

$$D = \min_{\omega\ \in\ \Omega} \mathbb{R}\text{ank}(\mathbf{H}_\omega).$$

An amplify-and-forward strategy utilizing only linear transformations at the relays (that do not depend on the channel realization) is sufficient to achieve this MMG.

Proof: (Outline) The proof proceeds as follows:

1) First, a converse for the MMG is provided using simple cut-set bounds.
2) Then, we convert the fading network into a deterministic network with the property that the cut-set bound on MMG for the fading network is the same as the cut-set bound on the capacity of the deterministic network.
3) We then characterize the zero-error capacity of the linear deterministic wireless network.
4) Finally, we convert a capacity-achieving scheme for the deterministic network into a MMG-achieving scheme for the fading network, which matches the converse.

■

The outline of the proof given above is detailed below. A converse on the degrees of freedom of a ss-ss fading network is immediate and is formalized in the following lemma.

Lemma 3.5: Given a ss-ss fading network with i.i.d. Rayleigh fading coefficients, the MMG, D, is upper bounded by the MMG of every cut:

$$D \quad \leq \quad \min_{\omega\ \in\ \Lambda} \mathbb{R}\text{ank}(\mathbf{H}_\omega), \tag{36}$$

where Λ denotes the set of all cuts in the network, and $\mathbf{H}_\omega$ is the matrix corresponding to the cut $\omega \in \Lambda$.

Next we proceed to the achievability part of the proof. First, we convert the wireless fading network into a derived linear deterministic network.[7] The construction of the derived deterministic network is described below. We will show that the zero-error capacity of this derived deterministic network is lower bounded by the upper-bound on the MMG of the fading network.

Let the number of edges in the fading network be N. Fading coefficients associated with edges of the network are denoted by $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_N$. To construct the derived deterministic network, consider a deterministic network with the same topology as that of the original fading network. We take q, the vector length in the deterministic network to be equal to the maximum number of antennas of any node in the fading network. For nodes with number of antennas less than q, we leave the remaining nodes unconnected. We still need to decide the finite field size, p, and finite field coefficients on all edges to completely characterize the equivalent finite-field deterministic network. We shall denote these finite-field coefficients by $\xi_i,\ i = 1, 2, \ldots, N$. We shall consider $\{\xi_i\}$ as indeterminates, before values are assigned to them.

For determining the field size p and $\{\xi_i,\ i = 1, 2, \ldots, N\}$, we will impose further conditions. In particular, we will ensure that the deterministic network will have at least the same capacity as the upper bound on MMG for the fading network. Due to the similarity between the expression for capacity in Theorem 3.2 and MMG terms in Lemma 3.5, above condition can be met by making sure that, cut-by-cut, the rank of the transfer matrix (G_ω) in the deterministic network is at least as large as the structural rank of the transfer matrix $\mathbf{H}_\omega$, i.e., $\text{rank}(G_\omega) \geq \mathbb{R}\text{ank}(\mathbf{H}_\omega)$.

Let us fix a cut ω, and let $r_\omega := \mathbb{R}\text{ank}(\mathbf{H}_\omega)$ be the structural rank of the transfer matrix of the cut in the fading network. Then, there exists a $r_\omega \times r_\omega$ sub-matrix (say $\mathbf{H}_\omega^{'}$) of $\mathbf{H}_\omega$, which has structural rank r_ω. Consider the same cut on the deterministic network and find the corresponding $r_\omega \times r_\omega$ sub-matrix $G_\omega^{'}$ of the transfer matrix G_ω. Now consider the determinant of the matrix $G_\omega^{'}$. The determinant is a polynomial in several variables $\xi_i, i = 1, 2, .., N$ with rational integer coefficients. Let us call this polynomial as $f_\omega(\xi_1, \xi_2, .., \xi_N)$. This polynomial is not identically a zero polynomial over $\mathbb{Q}$. This is because if it had been, then the substitution of $\xi_i = h_i$ will also yields zero irrespective of the choice of h_i, making the determinant zero even for the gaussian case, leading to a contradiction. Therefore, f_ω is a non-zero polynomial. We also observe that the degree of f_ω in each of the variable ξ_i is at most one. The lemma below, easily proved using elementary algebra, shows that it is possible to identify a finite field $\mathbb{F}_p$ and an allocation to $\{\xi_i\}$ with numbers from $\mathbb{F}_p$ such that f_ω does not vanish.

Lemma 3.6: Given a polynomial $f(\xi_1, \xi_2, ..., \xi_N)$ with integer coefficients, which is not identically zero, there exists a prime field $\mathbb{F}_p$ with p large enough, such that the polynomial evaluates to a non-zero value at least for one assignment of field values to the formal variables.

[7] It must be noted that the conversion to deterministic network used here is different from that used in [32] and [38].

However we want ensure the above condition for every cut in the network. To do so, consider the polynomial

$$f(\xi_1, \xi_2, .., \xi_N) := \prod_{\omega \in \Omega} f_\omega(\xi_1, \xi_2, .., \xi_N). \tag{37}$$

Now, the polynomial f is non-zero since it is a product of non-zero polynomials f_ω and the degree of f in any of the variables is at-most $|\Omega|$. We want a field $\mathbb{F}_p$ and an assignment for ξ_i from the field such that f is nonzero. Using Lemma 3.6, such an assignment exists. Let us choose that p and the assignment that makes f non-zero. Thus we have a deterministic wireless network whose capacity is guaranteed to be greater than or equal to the MMG upper bound, given in (36).

Next, we prove that the zero error capacity, C_{ZE}, of a linear deterministic network is equal to its ϵ-error capacity.

Definition 4: [39] The zero error capacity of a channel is defined as the supremum of all achievable rates across the channels such that the probability of error is exactly zero.

Theorem 3.7: The zero error capacity of a ss-ss deterministic wireless network is equal to

$$C_{ZE} \quad = \quad \min_{\omega \ \in \ \Omega} \mathrm{rank}(G_\omega)$$

This capacity can be achieved using a linear code and linear transformations in all relays.

Proof: We will prove this theorem using the ϵ-error capacity result from Theorem 3.2. Let the ss-ss deterministic network be composed of M relay nodes. From the achievability result in the proof of Theorem 3.2, given any $\epsilon > 0$ and rate $r < C$, there exists a block length T, number of blocks L, set of linear transformations $A_j, j = 1, 2, ..., M$ of size $qT \times qT$ used by all relays and a code book $\mathcal{C}$ for the source, such that the average probability of error, P_e, is less than or equal to ϵ. Each codeword $x_i \in \mathcal{C}$ is a $qT \times 1$ vector that specifies the entire transmission from the source. Let $x_1, ..., x_{|\mathcal{C}|}$ be the codewords.

Let us assume that the sink listens for $L^{'} \geq L$ blocks in general to account for the presence of paths of unequal lengths in the network between source and sink, (for large L, we would have $\frac{L^{'}}{L} \rightarrow 1$, so this does not affect rate calculations). Let $M := LT$ and $M^{'} := L^{'}T$. The transfer equation between the source and the destination vectors are specified by, $\mathbf{y} = G\mathbf{x}$ since all transformations in the network are linear. Here G is a $qM^{'} \times qM$ matrix, $\mathbf{x}$ is the M-length transmitted vector, and $\mathbf{y}$ is the $M^{'}$-length received vector.

Given the transmitted vector $\mathbf{x}_i$ corresponding to a message m_i at the source, the decoder either makes an error always or never makes an error. This is because the channel is a deterministic linear map, and error is only due to the fact that x_i and x_j are mapped to the same vector at the decoder. Let P_e^i be the probability of error, conditioned on the fact that the i-th codeword, $\mathbf{x}_i$ is transmitted. Then $P_e^i \in \{0, 1\}$ according to the argument above and the average codeword error probability

$$P_e = \frac{1}{|\mathcal{C}|} \sum_{i=1}^{|\mathcal{C}|} P_e^i \ \leq \ \epsilon \quad \Rightarrow \quad \sum_{i=1}^{|\mathcal{C}|} P_e^i \leq \epsilon |\mathcal{C}|.$$

This means that at least $(1-\epsilon)|\mathcal{C}|$ codewords have zero probability of error. Therefore if we choose only these $(1-\epsilon)|\mathcal{C}|$ codewords as an expurgated code $\mathcal{C}^{'}$, then the code has zero probability of error under the same relay matrices and decoding rule. The rate of the code is however $\bar{r} = r - \frac{\log\{(1-\epsilon)^{-1}\}}{M}$. Let $\delta = \frac{\log\{(1-\epsilon)^{-1}\}}{M}$ be the rate loss and therefore, the expurgated code has negligible rate loss as M becomes large. Now, we have established a zero error code of rate $r - \delta$. By choosing r arbitrarily close to C and M large, we get $C_{ZE} = C$.

The code $\mathcal{C}$, as given in [31] is non-linear, and so is the case of $\mathcal{C}^{'}$. However, we can obtain a linear code with zero probability of error. Since there exists a zero error code for rate $\bar{r}$ with block length T and number of blocks L, it means that the transfer matrix G between the source and the sink has rank at least $\bar{r}M$. Hence G has a sub-matrix $G^{'}$ of size $\bar{r}M \times \bar{r}M$, which is of full rank. By activating appropriate nodes, we can obtain the effective transfer matrix to be $G^{'}$. In that case, a linear code of rate $\bar{r}$ which communicates only on the $\bar{r}M$ subspaces can be used to achieve zero error. ■

Thus, for a given fading network, we have constructed an equivalent deterministic network. In the equivalent network, we also have a zero-error achievable rate $\bar{r}$ using a linear code of block length T and L blocks with linear strategies at the relays. This achievable rate is related to the MMG of the original fading network as follows:

$$\begin{aligned} \bar{r} + \delta \quad &= \quad C_{ZE} \\ &= \quad \min_{\omega \ \in \ \Omega} \mathrm{rank}(G_\omega) \\ &\geq \quad \min_{\omega \ \in \ \Omega} \mathbb{R}\mathrm{ank}(\mathbf{H}_\omega) \end{aligned}$$

Further, the positive constant δ can be made as small as possible as we wish by increasing the block length T. Now, when we use the zero-error scheme detailed above, the transfer matrix G of size $qM \times qM$ between the input and the output vectors $\mathbf{x}$ and $\mathbf{y}$ is at least of rank $\bar{r}M \approx CM$.

Finally, we lift the achievability strategy of zero-error capacity in the equivalent deterministic networks to arrive at an achievable strategy for MMG in corresponding fading network.

In the reduced deterministic network of a fading network, to achieve the zero-error capacity, the relays perform matrix operations A_i on received vectors for T time durations. Since each received vector is of size q, the matrix A_i is of size $qT \times qT$. Now we use the same strategy for the fading network, i.e., all relays use the same matrices A_i, that are obtained via the zero-error strategy in the reduced deterministic network. Though the entries of A_i belong to $\mathbb{F}_p$, they can be treated as integers by identifying the elements of $\mathbb{F}_p$ with the integers $0, 1, , ..., (p-1)$. Therefore the matrices A_i can also be interpreted as matrices over $\mathbb{C}$. By using linear maps A_i at relays in the fading network, we get an induced channel matrix $\mathbf{H}$, and effective channel would be of the form, $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. As is shown in Theorem 2.11, MMG offered by this channel is equal to $\mathbb{R}\mathrm{ank}(\mathbf{H})$. We shall prove that MMG offered by this induced channel is greater than or equal to $\bar{r}M$, i.e., to show that $\mathbb{R}\mathrm{ank}(\mathbf{H}) \geq \bar{r}M$. That is equivalent to show that

there exists an assignment of $\mathbf{h}_i = h_i$ in the fading network such that $\text{rank}(H) \geq \bar{r}M$.

In the proof of Theorem 3.7, we restricted the operation of the derived deterministic network to create a transfer matrix G of size $qM \times qM$ with rank greater than or equal to $\bar{r}M$. Now we have a similar transfer matrix $\mathbf{H}$ in the fading network. If we assign the underlying random variables $\mathbf{h}_i$ to be equal to ξ_i, again by identifying the elements of $\mathbb{F}_p$ with the integers $0, 1, , ..., (p-1)$, we have an assignment of $\mathbf{H}$ that has rank at least $\bar{r}M$. Since the strucutral rank is the maximum possible rank under any assignment, we get that,

$$\mathbb{R}\text{ank}(\mathbf{H}) \quad \geq \quad \text{rank}(G) \geq \bar{r}M.$$

The induced channel therefore has a MMG equal to $\bar{r}M$ by Theorem 2.11. Since the network is operated for $M = LT$ time slots in order to obtain a MMG greater than or equal to $\bar{r}M$, the MMG of the network per time slot is greater than or equal to $\bar{r}$. By increasing the block length T and the number of blocks L, the achievable MMG can be made arbitrarily close to C_{ZE}. Thus the upper bound given in Lemma 3.5 is achieved, and hence MMG of ss-ss fading network is given by

$$D \quad = \quad \min_{\{\omega \ \in \ \Lambda\}} \mathbb{R}\text{ank}(\mathbf{H}_\Lambda).$$

D. MMG for Multi-casting

In this section, we extend the result on MMG to the multi-casting scenario.

Theorem 3.8: Given a single-source D-sink multi-cast gaussian wireless network, with Rayleigh fading coefficients, the MMG of the network is given by

$$D = \min_{\{j=1,2,...,D\}} \ \min_{\omega \ \in \ \Omega_j} \mathbb{R}\text{ank}(\mathbf{H}_\omega). \tag{38}$$

An amplify-and-forward strategy utilizing only linear transformations at the relays is sufficient to achieve this MMG.

Proof: The proof uses Theorem 3.3, goes in the similar lines of that of Theorem 3.4 and is omitted here for brevity. ■

IV. DMT Bounds for Single Antenna Relay Networks

In this section, we consider ss-ss networks equipped with full-duplex single-antenna nodes. We provide a lower bound to the DMT of such a network by exploiting Menger's theorem.

Definition 5: Consider a network N and a path P from source to sink. This path P is said to have a *shortcut* if there is a single edge in N connecting two non-consecutive nodes in P.

Theorem 4.1: Consider a ss-ss full-duplex network with single antenna nodes. Let the min-cut of the network be M. Let the network satisfy *either* of the two conditions below:

1) The network has no directed cycles, or
2) There exist a set of M edge-disjoint paths between source and sink such that *none* of the M paths have shortcuts.

Then, a linear DMT $d(r) = M(1-r)^+$ between a maximum multiplexing gain of 1 and maximum diversity M is achievable.

Proof: Given that the network has min-cut M, there are M edge-disjoint paths from source to sink by Menger's theorem [51]. Let us label the edge-disjoint paths $e_1, e_2, ..., e_M$. Let the product of the fading coefficients along the path e_i be g_i. Let D_i be the delay of each path. Let $D = \max D_i$. We add delay $D - D_i$ to the path e_i such that all paths now are of equal delay. We activate the edges as follows:

1) Activate all edges in the edge-disjoint path e_1 simultaneously for a period T, where $T >> D$. The source, on the first $T - D$ activations, will transmit $(T - D)$ coded information symbols, followed by a sequence of D zero symbols. The reason for this will become clear shortly. The net effect will be to create a $((T - D) \times (T - D)$ transfer matrix $\mathbf{H}_1$ from the $(T - D)$ source symbols to the last $(T - D)$ symbols received by the destination (the first D symbols received by the destination are all zero).
   The matrix $\mathbf{H}_1$ will be either upper-triangular or lower-triangular, with the elements along the diagonal all equal to the path gain $\mathbf{g}_1$ on path e_1, according to whether the condition 1 or condition 2 of the theorem is satified. First, we explain the case when the graph has no directed cycles. In this case, off-diagonal terms above the diagonal can arise due to the presence of short-cuts. However, off-diagonal terms below the diagonal would constitute a directed cycle and will thus not appear. Therefore the matrix will be upper-triangular in this case. Next, for the case when the graph has no shortcuts, off-diagonal terms below the diagonal can arise due to the presence of cycles in the graph. However, no terms above the diagonal will be present because of the presumed absence of shortcuts. Thus the induced matrix will be lower-triangular in structure.
2) Repeat *Step 1* for all edge-disjoint paths $e_2, ..., e_M$. The net transfer matrix $\mathbf{H}$ will be block diagonal of the form

$$\mathbf{H} \quad = \quad \begin{bmatrix} \mathbf{H}_1 & & & \\ & \mathbf{H}_2 & & \\ & & \ddots & \\ & & & \mathbf{H}_M \end{bmatrix}. \tag{39}$$

   composed of M blocks along the diagonal, one corresponding to each path, and either all of them are upper triangular or all are lower triangular by the argument above.

Now, if $d(r)$ is the DMT of the network operating under the protocol given above, then

$$d(r) \quad = \quad d_H(MTr), \tag{40}$$

since MT time instants were used up by the protocol in order to obtain the indcued channel matrix $\mathbf{H}$. We next proceed to lower bound $d_H(r)$. By Theorem 2.12, we have the lower bound

$$d_H(r) \quad \geq \quad d_{H^{(0)}}(r) \tag{41}$$

where $\mathbf{H}^{(0)}$ is the matrix corresponding to the diagonal terms in $\mathbf{H}$. Next, we observe that $\mathbf{H}^{(0)}$ corresponds to a parallel channel with M fading coefficients $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_M$, each of them repeated $T-D$ times. We can compute the DMT of this parallel channel using Lemma 2.8. Thus we get,

$$d_{H^{(0)}}(r) = M\left(1 - \frac{r}{M(T-D)}\right)^{+} \tag{42}$$
$$\Rightarrow d(r) = d_H(MTr) \geq d_{H^{(0)}}(MTr) \tag{43}$$

$$d(r) = M\left(1 - \frac{MTr}{M(T-D)}\right)^{+}. \tag{44}$$

For T tending to ∞, we get $d(r) \geq M(1-r)^{+}$. ∎

Appendix I
Proof of Lemma 2.2

Let the multinomial f be written as a sum of S monomials $f(X_1, X_2, ..., X_M) := \sum_{i=1}^{S} c_i f_i(X_1, X_2, ..., X_M)$, where for every i, c_i is a constant and f_i is a monomial, i.e., is comprised only of product of powers of X_i. Then for every assignment, $X_i = h_i$,

$$|f(h_1, h_2, ..., h_M)| \leq \sum_i |c_i||f_i(h_1, h_2, ..., h_M)|.$$

Now we have,

$$\begin{aligned}
&\Pr\{|f(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > k\} \\
&= \Pr\{|f(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)| > k^{\frac{1}{2}}\} \\
&\leq \Pr\{\sum_i |c_i||f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)| > k^{\frac{1}{2}}\} \qquad (45) \\
&\leq \Pr\left\{\bigcup_i \left(|c_i||f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)| > \frac{k^{\frac{1}{2}}}{S}\right)\right\} \\
&\leq \sum_i \Pr\left\{|c_i|^2|f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > \frac{k}{S^2}\right\} \\
&\leq \sum_i \Pr\left\{|f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > \frac{k}{c_{max}S^2}\right\}, \qquad (46)
\end{aligned}$$

where c_{max} is the maximum over all $\{|c_i|^2\}$. Now $|f_i(h_1, h_2, ..., h_M)|^2$ is a monomial in $|h_i|^2$ as well. Define, $\mathbf{u}_j := |\mathbf{h}_j|^2$. Then $\mathbf{u}_j$ is the squared norm of a $\mathbb{CN}(0,1)$ random variable $\mathbf{h}_j$, and therefore has an exponential distribution. We will regard $|f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2$ as a monomial $g_i(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_M)$ in $\{\mathbf{u}_j\}$. Thus

$$g_i(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M) = \prod_{j=1}^{M} \mathbf{u}_j{}^{a_{ij}},$$

where $0 \leq a_{ij} \leq D$ is an integer, where D is the maximum degree of any of the monomials g_i in any of the variables $\mathbf{h}_i$.

Now gathering a single term in the summation in RHS of (46), we have,

$$\begin{aligned}
&\Pr\left\{|f_i(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > \frac{k}{c_{max}S^2}\right\} \\
&= \Pr\left\{g_i(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M) > \frac{k}{c_{max}S^2}\right\} \\
&\leq \Pr\left\{\bigcup_j \left(\mathbf{u}_j{}^{a_{ij}} > \left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{M}}\right)\right\} \\
&\leq \sum_j \Pr\left\{\mathbf{u}_j{}^{a_{ij}} > \left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{M}}\right\} \\
&= \sum_j \Pr\left\{\mathbf{u}_j > \left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{Ma_{ij}}}\right\} \\
&\leq \sum_j \Pr\left\{\mathbf{u}_j > \left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{MD}}\right\} \qquad (47) \\
&= M \ \Pr\left\{\mathbf{u}_j > \left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{MD}}\right\} \\
&= M \ \exp\left(-\left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{MD}}\right). \qquad (48)
\end{aligned}$$

Equation (47) follows if

$$k \geq c_{max}S^2.$$

We get this condition by setting $\delta := c_{max}S^2 > 0$, since by the hypothesis of the lemma, we have $k \geq \delta$.

Combining (46) and (48), we get, for $k \geq \delta$,

$$\begin{aligned}
&\Pr\{|f(\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M)|^2 > k\} \\
&\leq MS \ \exp\left(-\left(\frac{k}{c_{max}S^2}\right)^{\frac{1}{MD}}\right) \\
&= A \ \exp\left(-Bk^{\frac{1}{d}}\right)
\end{aligned}$$

where $A := MS$, $B := \left(\frac{1}{c_{max}S^2}\right)^{\frac{1}{MD}}$, $\delta = c_{max}S^2$ and $d := MD$.

Appendix II
Proof of Theorem 2.3

$$\begin{aligned}
&Pr(\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}\Sigma^{-1}) \leq r\log\rho) \\
&\doteq Pr(\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}) \leq r\log\rho).
\end{aligned}$$

Let the correlation matrix of the noise vector be denoted by $\mathbf{\Sigma}$. The noise covariance matrix $\mathbf{\Sigma}$ depends on the channel realization and is therefore a random matrix, given by,

$$\begin{aligned}
\mathbf{\Sigma} &= \mathbb{E}_{\mathbf{z}}[\mathbf{z}\mathbf{z}^{\dagger}] \\
&= I + \sum_{i=1}^{M} \mathbf{G}_i\mathbf{G}_i^{\dagger}. \qquad (49)
\end{aligned}$$

Let $\lambda_i(A)$, $\lambda_{max}(A)$ and $\lambda_{min}(A)$ denote the i^{th} largest, maximum and minimum eigenvalues of a positive semi-definite

matrix A. If the context is clear, we may avoid specifying the matrix, and just use λ_i, λ_{max} and λ_{min} respectively.

To prove the lemma, we will use the Amir-Moez bound on the eigen values of the product of Hermitian, positive-definite matrices [52]. By this bound, for any two positive definite $n \times n$ Hermitian matrices A, B:

$$\lambda_i(A)\lambda_{min}(B) \leq \quad \lambda_i(AB) \quad \leq \lambda_i(A)\lambda_{max}(B).$$

So we get,

$$\begin{aligned}\det(I+\rho AB) &= \prod_i (1+\rho\lambda_i(AB)) \\ &\leq \prod_i (1+\rho\lambda_i(A)\lambda_{\max}(B)) \\ &= \det(I+\rho\lambda_{\max}(B)A).\end{aligned}$$

Similarly,

$$\det(I+\rho AB) \quad \geq \quad \det(I+\rho\lambda_{\min}(B)A).$$

Therefore, for any two positive definite $n \times n$ Hermitian matrices A, B,

$$\begin{aligned}&\det(I+\rho\lambda_{\min}(B)A) \\ &\quad \leq \det(I+\rho AB) \leq \det(I+\rho\lambda_{\max}(B)A). \quad (50)\end{aligned}$$

Applying (50) to $A = \mathbf{H}\mathbf{H}^\dagger$ and $B = \mathbf{\Sigma}^{-1}$, we get

$$\begin{aligned}&\Rightarrow \det(I+\rho\mathbf{H}\mathbf{H}^\dagger\lambda_{min}(\mathbf{\Sigma}^{-1})) \\ &\qquad \leq \det(I+\rho\mathbf{H}\mathbf{H}^\dagger\mathbf{\Sigma}^{-1}) \quad (51) \\ &\qquad \leq \det(I+\rho\mathbf{H}\mathbf{H}^\dagger\lambda_{max}(\mathbf{\Sigma}^{-1})). \quad (52)\end{aligned}$$

Continuing from (51) and (52), we have

$$\begin{aligned}&\Pr\{\log(\det(I+\rho HH^\dagger\lambda_{min}(\mathbf{\Sigma}^{-1}))) < r\log\rho\} \\ &\quad \geq \Pr\{\log\det(I+\rho HH^\dagger\mathbf{\Sigma}^{-1}) < r\log\rho\} \\ &\quad \geq \Pr\{\log(\det(I+\rho HH^\dagger\lambda_{max}(\mathbf{\Sigma}^{-1}))) < r\log\rho\}. \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (53)\end{aligned}$$

In the following, we will prove that both the bounds coincide as $\rho \to \infty$. We begin with the bounds on $\lambda_{min}(\mathbf{\Sigma})$ and $\lambda_{max}(\mathbf{\Sigma})$. In order to show that the lower and the upper bounds on the expression converge to the value $\Pr\{\log(\det(I+\rho HH^\dagger)) < r\log\rho\}$, we need to provide a lower bound for each $\lambda_i(\mathbf{\Sigma})$. Let e_i be the eigen vector corresponding to $\lambda_i(\Sigma)$ for every realization Σ of $\mathbf{\Sigma}$. Then,

$$\begin{aligned}\lambda_i(\Sigma) \parallel e_i \parallel^2 &= {e_i}^\dagger \Sigma e_i \\ &= {e_i}^\dagger \left(I + \sum_{i=1}^{M} \mathbf{G}_i\mathbf{G}_i^\dagger\right) e_i \\ &= \parallel e_i \parallel^2 + {e_i}^\dagger \left(\sum_{i=1}^{M} \mathbf{G}_i\mathbf{G}_i^\dagger\right) e_i \\ &\geq \parallel e_i \parallel^2 \\ \Rightarrow \lambda_i(\mathbf{\Sigma}) &\geq 1 \quad \forall i \\ \Rightarrow \lambda_{min}(\mathbf{\Sigma}) &\geq 1 \qquad\qquad (54) \\ \Rightarrow \lambda_{max}(\mathbf{\Sigma^{-1}}) &\leq 1\end{aligned}$$

Hence,

$$\begin{aligned}&\Pr\{\log\det(I+\rho\mathbf{H}\mathbf{H}^\dagger\mathbf{\Sigma}^{-1}) < r\log\rho\} \\ &\geq \Pr\{\log\det(I+\rho\mathbf{H}\mathbf{H}^\dagger\lambda_{max}(\mathbf{\Sigma}^{-1})) < r\log\rho\} \\ &\geq \Pr\{\log\det(I+\rho\mathbf{H}\mathbf{H}^\dagger) < r\log\rho\} \quad (55) \\ &\doteq \rho^{d_{out}(r)}. \quad (56)\end{aligned}$$

Now we proceed to get an upper bound on $\lambda_{max}(\mathbf{\Sigma})$:

$$\begin{aligned}\lambda_{max}(\mathbf{\Sigma}) &= \lambda_{max}(I + \sum_{i=1}^{M}\mathbf{G}_i\mathbf{G}_i^\dagger) \\ &= 1 + \lambda_{max}(\sum_{i=1}^{M}\mathbf{G}_i\mathbf{G}_i^\dagger)\end{aligned}$$

i.e.,

$$\begin{aligned}\lambda_{max}(\mathbf{\Sigma}) &\leq 1 + \text{Tr }(\sum_{i=1}^{M}\mathbf{G}_i\mathbf{G}_i^\dagger) \\ &= 1 + \sum_{i=1}^{M}\text{Tr }(\mathbf{G}_i\mathbf{G}_i^\dagger) \\ &= 1 + \sum_{i=1}^{M}||\mathbf{G}_i||^2{}_F \\ &= 1 + \sum_{i=1}^{M}\sum_{j=1}^{N^2}|f_{ij}|^2, \quad (57)\end{aligned}$$

where f_{ij} represents a polynomial entry of the matrix G_i. Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{2L}$ denote in some order, the real and imaginary parts of $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_L$. Then, the right hand side in (57) is a polynomial in the variables, $\mathbf{x}_i, i = 1, 2, ..., 2L$.

This leads to the following inequality,

$$\lambda_{max}(\mathbf{\Sigma}) \leq g(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2L}) + 1, \quad (58)$$

where $g(x_1, x_2, \ldots, x_{2L})$ is a polynomial without constant term in the variables $\{x_i\}$. Let us invoke Lemma 2.2 for the polynomial g which does not possess any constant term. The lemma is valid for all $k \geq \delta$, where δ depends on g. Let us choose ρ_0 such that $\rho_0^\epsilon - 1 \geq \delta$ and therefore $\forall \rho > \rho_0$, we have that $\rho^\epsilon - 1 \geq \delta$. Now, $\forall \rho > \rho_0$, we can invoke Lemma 2.2 and get,

$$\begin{aligned}\Pr\{\lambda_{max}(\mathbf{\Sigma}) > \rho^\epsilon\} &\leq \Pr\{g(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_S) > \rho^\epsilon - 1\} \\ &\leq A\exp(-B(\rho^\epsilon - 1)^{\frac{1}{d}}), \quad (59)\end{aligned}$$

for some constants A, B, $d > 0$.

Let $\mathcal{H}$ denote the set of all the fading coefficients in the network, and let $\hbar \in \mathcal{H}$ denote a realization of the fading coefficients. Thus $\hbar$ will be a vector specifying $h_i, i = 1, 2, ..., L$. Clearly, once a $\hbar$ is given, the values of the matrices H and G_i are all well defined, since all of them depend only on h_i.

Let $A = \{\hbar \in \mathcal{H} \mid \log\det(I+\rho HH^\dagger\Sigma^{-1}) < \rho^r\}$ and

$B = \{\hbar \in \mathcal{H} \mid \lambda_{max}(\Sigma) > \rho^{\epsilon}\}$ be two events. Then,

$$\begin{aligned}
\Pr(A) &= \Pr(A \cap B^c) + \Pr(A \cap B) \\
&\leq \Pr(A \cap B^c) + \Pr(B) \\
&\dot{\leq} \Pr(A \cap B^c) + A\exp(-B(\rho^{\epsilon} - 1)^{\frac{1}{d}}) \qquad (60) \\
\text{Now, } A &= \{\hbar \in \mathcal{H} \mid \log\det(I + \rho H H^{\dagger}\Sigma^{-1}) < \rho^r\} \\
&\subset \{\hbar \in \mathcal{H} \mid \log\det(I + \rho H H^{\dagger}\lambda_{min}(\Sigma^{-1})) < \rho^r\} \\
&= \{\hbar \in \mathcal{H} \mid \log\det(I + \rho H H^{\dagger}(\lambda_{max}(\Sigma))^{-1}) < \rho^r\} \\
A \cap B^c &\subset \{\hbar \in \mathcal{H} \mid \log\det(I + \rho^{1-\epsilon} H H^{\dagger}) < \rho^r\}. \qquad (61)
\end{aligned}$$

Substituting (61) and (59) in (60), taking logarithms and dividing by $\log\rho$ on both sides, we have,

$$\begin{aligned}
&\frac{\log \Pr(A)}{\log\rho} \\
&\leq \frac{\log[\Pr(A \cap B^c) + \Pr(B)]}{\log\rho} \\
&\dot{\leq} \frac{1}{\log\rho} \cdot \log\ [\ \Pr\{\hbar \in \mathcal{H} \mid \log\det(I + \rho^{1-\epsilon} H H^{\dagger}) < \rho^r\} \\
&\qquad + A\exp(-B(\rho^{\epsilon} - 1)^{\frac{1}{d}})\ ]. \qquad (62)
\end{aligned}$$

$$\begin{aligned}
&\lim_{\rho\to\infty} \frac{\log \Pr(A)}{\log\rho} \\
&\dot{\leq} \lim_{\rho\to\infty} \frac{\log[\Pr\{\hbar \in \mathcal{H} \mid \log\det(I + \rho^{1-\epsilon} H H^{\dagger}) < \rho^r\}]}{\log\rho}.
\end{aligned}$$

The last equation follows from (62), since the first term in the RHS of (62) varies inversely with an exponent of ρ whereas the second term is exponential in ρ, and therefore the sum is dominated by the first term. After making the variable change, $\rho^{'} = \rho^{1-\epsilon}$ and replacing ρ by $\rho^{'}$, we get

$$\begin{aligned}
&\Rightarrow \lim_{\rho\to\infty} \frac{\log \Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}\mathbf{\Sigma}^{-1}) < \rho^r\}}{\log\rho} \\
&\dot{\leq} (1-\epsilon) \\
&\qquad \lim_{\rho\to\infty} \frac{\log \Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}) < \rho^{(\frac{r}{1-\epsilon})}\}}{\log\rho} \\
&\doteq (1-\epsilon)\rho^{d_{out}(\frac{r}{1-\epsilon})}. \qquad (63)
\end{aligned}$$

In (63), ϵ is arbitrary, and we let it tend to zero. Hence, by (63) and (56), the exponents for both the bounds in (53) coincide and we obtain,

$$\begin{aligned}
&\Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}\Sigma^{-1}) < r\log\rho\} \\
&\doteq \Pr\{\log\det(I + \rho\mathbf{H}\mathbf{H}^{\dagger}) < r\log\rho\}.
\end{aligned}$$

This proves the assertion of the theorem.

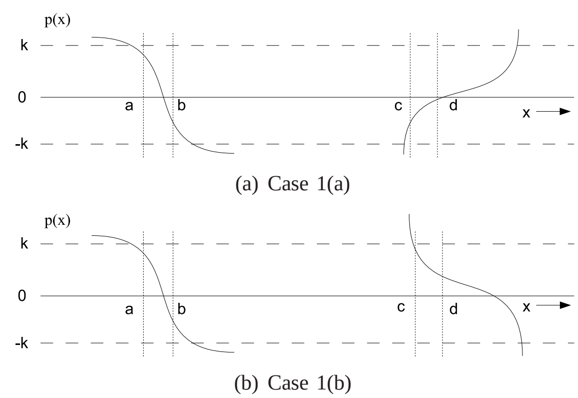
Appendix III
Proof of Lemma 2.9

Consider any two intervals $R_1, R_2 \in \mathcal{R}$. If we are not able to find two such intervals, then clearly $L \leq 1$, and we are done. Let $R_1 = [a, b]$ and $R_2 = [c, d]$, and without loss of generality assume that $a < b < c < d$, since they are, by hypothesis, disjoint. First, we claim that there exists a point $b \leq x_0 \leq c$, such that either $p'(x_0) = 0$, or $p''(x_0) = 0$. We now proceed to prove this claim.

Clearly, either of the two conditions (10) or (11) is violated just to the right of the point $x = b$, else, the interval would extend beyond b. We consider two cases.

Case 1: $|p(b)| \geq k$.

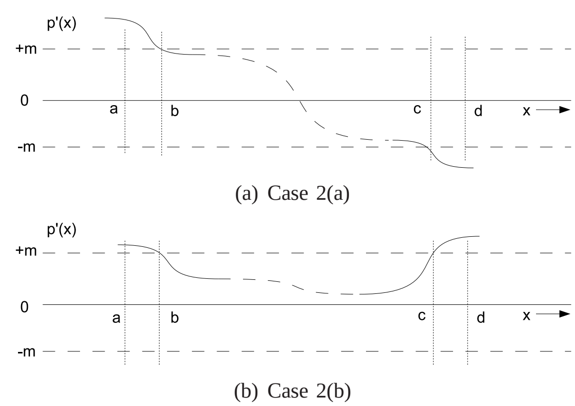
Condition (10) is violated in the region in the immediate right of the interval $[a, b]$. This implies that the absolute value of the evaluation of polynomial function, $|p(x)|$, has to be greater than k in the beginning of the interval $[b, c]$. Also, we know that within $[a, b]$ and $[c, d]$, $|p(x)|$ is strictly less than k. This can happen in two ways, as shown in Fig. 7 (the other possibility is that the polynomial can be the negative of that shown in the figure, in which case the same argument holds). In either of these ways, the function has to go through $-k$ value twice in $[b, c]$. Therefore, by Rolle's theorem, $p'(x_0) = 0$, for some $b \leq x_0 \leq c$.



(a) Case 1(a)

(b) Case 1(b)

Fig. 7. Condition (10) is violated in $[b, c]$



(a) Case 2(a)

(b) Case 2(b)

Fig. 8. Condition (11) is violated in $[b, c]$.

Case 2: Condition (11) is violated at the end of interval $[a, b]$

This implies that the absolute value of the evaluation of the derivative of $p(x)$, i.e.,$|p'(x)|$, diminishes below m in the beginning of the interval $[b, c]$. Also, we know that within $[a, b]$ and $[c, d]$, $|p'(x)|$ is greater than or equal to m. This can happen only in two ways, as shown in Fig. 8. In the first case, $p'(x_0) = 0$, for some $b \leq x_0 \leq c$. In the second case, the function $p'(x)$ takes the same value $+m$ twice in $[b, c]$, and hence by Rolle's theorem, $p''(x_0) = 0$, for some $b \leq x_0 \leq c$.

By the above claim, for any two arbitrary intervals in $\mathcal{R}$, there exists real root of $p'(x)$ or $p''(x)$ between those

two intervals. Since the number of roots of a polynomial is bounded by its degree, there will be only finitely many such intervals. In particular, the number of intervals L is bounded by $2d$, which is an upper bound on the total number of zeros of $p(x)$ and $p^{'}(x)$.

Appendix IV
Proof of Lemma 2.10

For any polynomial f in several gaussian random variables, we have that

$$\Pr\{f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \neq 0\} = 1.$$

This follows since letting

$$\begin{aligned} \underline{y} &= (x_1, x_2, \cdots, x_{N-1}) \\ S &= \{x_N \mid f(x_1, x_2, \cdots, x_N) = 0\}, \end{aligned}$$

we see that

$$\begin{aligned} \Pr\{f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = 0\} &= \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\underline{y}) \int_{x_N \in S} f(x_N/\underline{y}) dx_N d\underline{y} &= 0, \end{aligned}$$

because the innermost integral equals zero as S is finite given a particular assignment of $\underline{\mathrm{y}}$, i.e.,

$$\int_{x_N \in S} f(x_N/\underline{y}) dx_N = 0. \tag{64}$$

Let $\mathbf{x} := \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$. Let us define an indicator function $I_\delta(\mathbf{x})$ as follows:

$$I_\delta(\mathbf{x}) := \left\{ \begin{array}{ll} 1, & |f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)| < \delta \\ 0, & \text{else} \end{array} \right\}.$$

Then

$$\begin{aligned} &\Pr\{|f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)|^2 < \delta\} &(65)\\ &= \mathbb{E}_{\mathbf{x}}\ I_k(\mathbf{x}) \\ &= \mathbb{E}_{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N-1}}\ \mathbb{E}_{\mathbf{x}_N} \{I_k(\mathbf{x}) \mid \mathbf{x}_1^{N-1}\} \\ &= \mathbb{E}_{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N-1}}\ \Pr\ \{|f(x_1, x_2, ..., x_{N-1}, \mathbf{x}_N)|^2 < \delta\}. &(66) \end{aligned}$$

Let $f(\mathbf{x_N}) := f(x_1, x_2, ..., x_{N-1}, \mathbf{x}_N)$, where the dependence of f on the first $N-1$ variables is made implicit. Let

$$f(\mathbf{x}_N) = \sum_{k=0}^{d_N} b_k \mathbf{x}_N^k,$$

where d_N is the degree of the polynomial f in the variable x_N. Since b_{d_N} is a polynomial in the variables x_1^{N-1}, it follows from the lemma above that with probability one, $b_{d_N} \neq 0$.

Let

$$g(x_N) = \frac{\partial f(x_N)}{\partial x_N}$$

be the partial derivative of $f(x_N)$ with respect to x_N. Then we can write

$$\begin{aligned} &\Pr\ \{|f(x_1, x_2, ..., x_{N-1}, \mathbf{x}_N)| < \delta\} &(67)\\ &= \Pr\ \{|f(\mathbf{x}_N)| < \delta, |g(\mathbf{x}_N)| \geq \delta^{1/2}\} \\ &\quad + \Pr\ \{|f(\mathbf{x}_N)| < \delta, |g(\mathbf{x}_N)| < \delta^{1/2}\} \\ &\leq \Pr\ \{|f(\mathbf{x}_N)| < \delta, |g(\mathbf{x}_N)| \geq \delta^{1/2}\} \\ &\quad + \Pr\ \{|g(\mathbf{x}_N)| < \delta^{1/2}\}. &(68) \end{aligned}$$

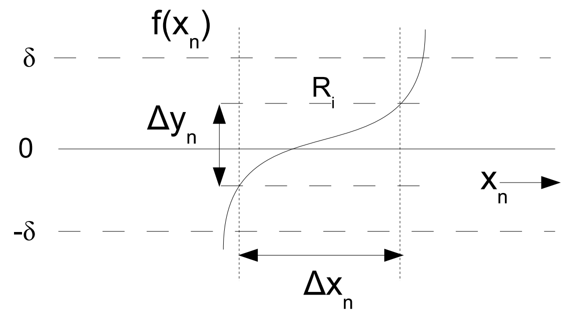


Fig. 9. f(x) in a region R_i

Let us consider the first term on the RHS. The region $\mathcal{R} := \{|f(x_N)| < \delta, |g(x_N)| \geq \delta^{1/2}\}$ is described by two conditions $|f(x_N)| < \delta$ and $|g(x_N)| \geq \delta^{1/2}$. It is shown in Lemma 2.9 that the set of all values of x_N satisfying both conditions can be expressed as the union of L pairwise-disjoint intervals $R_i, i = 1, 2, ..., L$ with $L < 2d_N$. Now $\Pr(x_N \in R) = \sum_{i=1}^{L} \Pr(\mathbf{x}_N \in R_i)$. We will now proceed to upper-bound the probability $\mathbf{x}_N \in R_i$. To do so, consider Fig. 9. Let Δx_n be the width of the interval R_i and $\Delta f(x_n)$ be the height (equal to the difference in maximum and minimum values of $f(.)$ in R_i). Since the slope of the curve $g(x)$ is greater than $\delta^{1/2}$ throughout R_i, we have that

$$\left| \frac{\Delta f(x_n)}{\Delta x_n} \right| \geq \delta^{1/2}.$$

For our purposes, we can assume without loss of generality that

$$\frac{\Delta f(x_n)}{\Delta x_n} \geq \delta^{1/2},$$

which gives us

$$\Delta x_n \leq \frac{\Delta f(x_n)}{\delta^{1/2}}.$$

However in any contiguous region, $\Delta f(x_n) \leq 2\delta$. This implies that

$$x_n \leq \frac{2\delta}{\delta^{1/2}} = 2\delta^{1/2}.$$

Since $\mathbf{x}_n$ is a $\mathcal{N}(0,1)$ random variable, we have that $\Pr\{\Delta \mathbf{x}_n \leq a\} \leq ca$, where c is the maximum value of the gaussian pdf.

Therefore,

$$\begin{aligned}\{\mathbf{x} \in R_i\} &\subset \{\mathbf{x}_n \le 2\delta^{1/2}\}\\ \Pr\{\mathbf{x} \in R_i\} &\le \Pr\{\mathbf{x}_n \le 2\delta^{1/2}\}\\ &\le 2c\delta^{1/2}.\end{aligned}$$

Using

$$\Pr\{\mathbf{x} \in R\} = \sum_{i=1}^{L} \Pr\{\mathbf{x} \in R_i\},$$

we obtain

$$\Pr\ \{|f(\mathbf{x}_N)| < \delta, |g(\mathbf{x}_N)| \ge \delta^{1/2}\} \le L2c\delta^{1/2} = C\delta^{1/2}. \tag{69}$$

Plugging (69) into (68) yields

$$\Pr\ \{|f(\mathbf{x}_N)| < \delta\} \le C\delta^{1/2} + \Pr\ \{|g(\mathbf{x}_N)| < \delta^{1/2}\}. \tag{70}$$

Since $g(x)$ is of lower degree than $f(x)$, the process can be continued to yield

$$\begin{aligned}\Pr\ \{|f(\mathbf{x}_N)| < \delta\} &\le C\delta^{1/2} + C\delta^{1/4} + ... + C\delta^{1/2^{d_N - 1}}\\ &\quad + \Pr\{d_N!\ b_{d_N} \le \delta^{1/2^{d_N}}\}.\end{aligned} \tag{71}$$

Only the last term involving b_{d_N} is a function of the remaining variables $x_1, x_2, ..., x_{N-1}$. We next substitute (71) into (66) and take the expectation over the remaining variables. This yields

$$\Pr\{|f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)| < \delta\} \tag{72}$$

$$\begin{aligned}&\le C\delta^{1/2} + C\delta^{1/4} + ... + C\delta^{1/2^{d_N-1}}\\ &\quad + \mathbb{E}_{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_{N-1}}\ I_{\{d_N!\ \mathbf{b}_{d_N} \le \delta^{1/2^{d_N}}\}}.\end{aligned} \tag{73}$$

The last term is identical to that in the right hand side of (66), except that the polynomial $d_N!\ b_{d_N}$ involves $(N-1)$ or fewer variables and hence this procedure can be continued. Eventually, we will be left with the probability that a constant coefficient J is greater than $\delta^{1/s}$ for some integer s. Choosing the constant K appearing in the statement of the lemma to equal J^s, we obtain that this probability is equal to the probability that $K \le \delta$. But by hypotheses, $K > \delta$ and hence this probability is equal to zero. This allows us to rewrite the bound on probability appearing in (73) as

$$\begin{aligned}&\Pr\{|f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)| < \delta\}\\ &\le C_1(\delta^{1/2} + \delta^{1/4} + ... + \delta^{1/2^e})\end{aligned}$$

for a suitable constant C_1 and some integer e.

Choosing $K \le 1$ forces $\delta < 1$ since by hypotheses, $\delta < K$. In this case

$$\delta^{1/2} + \delta^{1/4} + ... + \delta^{1/2^e} \le e\delta^{1/2^e}$$

With $A := eC_1$ and $d := 2^e$, we get

$$\Pr\{|f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)| < \delta\} \le A\delta^{1/d} \tag{74}$$

as desired.

APPENDIX V
PROOF OF LEMMA 2.5

Let $\mathbf{x}, \mathbf{y}, \mathbf{H}$ denote the concatenated input and output vectors and channel matrix respectively, i.e., $\mathbf{x} = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}]^T, \mathbf{y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_M}]^T$ and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & & & \\ & \mathbf{H}_2 & & \\ & & \ddots & \\ & & & \mathbf{H}_M \end{bmatrix}. \tag{75}$$

Then the input-output relation of the parallel channel is given by $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. We now proceed to determine the probability of outage. We have:

$$\begin{aligned}&I(\mathbf{x}; \mathbf{y}|\mathbf{H} = H)\\ &= h(\mathbf{y}|\mathbf{H} = H) - \sum_{i=1}^{M} h(\mathbf{y}_i|\mathbf{y}_1^{i-1}, \mathbf{x}, \mathbf{H} = H)\\ &= h(\mathbf{y}|\mathbf{H} = H) - \sum_{i=1}^{M} h(\mathbf{y}_i|\mathbf{x}_i, \mathbf{H} = H)\\ &\le \sum_{i=1}^{M} h(\mathbf{y}_i|\mathbf{H} = H) - \sum_{i=1}^{M} h(\mathbf{y}_i|\mathbf{x}_i, \mathbf{H} = H)\\ &= \sum_{i=1}^{M} [h(\mathbf{y}_i|\mathbf{H} = H) - h(\mathbf{y}_i|\mathbf{x}_i, \mathbf{H} = H)]\\ &= \sum_{i=1}^{M} [h(\mathbf{y}_i|\mathbf{H}_i = H_i) - h(\mathbf{y}_i|\mathbf{x}_i, \mathbf{H}_i = H_i)]\\ &= \sum_{i=1}^{M} I(\mathbf{x}_i; \mathbf{y}_i|\mathbf{H}_i = H_i).\end{aligned} \tag{76}$$

$$\begin{aligned}&\Rightarrow \Pr\{I(\mathbf{x}; \mathbf{y} \mid \mathbf{H} = H) \le r \log \rho\}\\ &\quad \ge \Pr\{\sum_{i=1}^{M} I(\mathbf{x}_i; \mathbf{y}_i|\mathbf{H}_i = H_i) \le r \log \rho\}\end{aligned}$$

Equality in the equation above holds if all the $\mathbf{x}_i$ are independent. So we will choose the $\mathbf{x}_i$ to be independent, for the rest of the discussion, since this maximizes the mutual information and hence minimizes the outage probability. Define $\mathbf{Z}_i := I(\mathbf{x}_i; \mathbf{y}_i|\mathbf{H}_i = H_i)$. Thus $\mathbf{Z}_i$ is a function of the channel realization H_i and is therefore a random variable. Since $\{\mathbf{H}_i\}$ are independent by the hypothesis of the lemma and $\mathbf{x}_i$ are independent by the argument above, $\{\mathbf{Z}_i\}$ are also independent.

Let $R = r\ \log(\rho)$ and $R_i = r_i\ \log(\rho)$ for $i = 1, 2, \ldots, M$. Our next goal is to evaluate $\Pr\{\ \sum_{i=1}^{M} \mathbf{Z}_i \le r \log(\rho)\ \}$. To do this, we first consider the case when $M = 2$ and we evaluate $\Pr\{\ \mathbf{Z}_1 + \mathbf{Z}_2 \le r \log(\rho)\ \}$. Then we extend this to general M

by induction. We define

$$
\begin{aligned}
F_{Z_i}(R_i) &:= \Pr\{\mathbf{Z}_i < R_i\} \\
f_{Z_i}(R_i) &:= \frac{d}{dR_i} F_{Z_i}(R_i) \\
\text{Let } F_{Z_i}(R_i) &\doteq \rho^{-d_i(r_i)} \\
\text{Then } f_{Z_i}(R_i) &\doteq \frac{d}{dr_i \ \log(\rho)} \rho^{-d_i(r_i)} \\
&\doteq \rho^{-d_i(r_i)}(-1)\frac{d}{dr_i} d_i(r_i) \\
&\doteq \rho^{-d_i(r_i)}.
\end{aligned}
$$

The last equation follows since $d_i(r_i)$ is a decreasing function making $-\frac{d}{dr_i}$ positive.

$$
\begin{aligned}
\Pr(\mathbf{Z}_1 + \mathbf{Z}_2 \le R) &= \rho^{-d(r)} \\
&= \int_0^\infty f_{Z_1}(R_1) F_{Z_2}(R - R_1) dR_1 \\
&\doteq \int_0^\infty \rho^{-d_1(r_1)} \ \rho^{-d_2(r-r_1)} \log(\rho) d(r_1).
\end{aligned}
$$

By Varadhan's Lemma [47], the SNR exponent of the integral is given by

$$
\begin{aligned}
d(r) &= \inf_{r_1 \ge 0} d_1(r_1) + d_2(r - r_1) \\
&= \inf_{(r_1, r_2):\ r_1 + r_2 = r} \ \sum_{i=1}^{2} d_i(r_i).
\end{aligned}
$$

Proceeding by induction, we get for the general case with M parallel channels where

$$
\rho^{-d(r)} \ \doteq \ \Pr\{ \sum_{i=1}^{M} Z_i \le r \log(\rho) \ \}
$$

that

$$
d(r) = \inf_{(r_1, r_2, \cdots, r_M):\ \sum_{i=1}^{M} r_i = r} \ \sum_{i=1}^{M} d_i(r_i).
$$

Appendix VI
Proof of Lemma 2.8

Following the same line of reasoning as in the proof of Lemma 2.5, we choose $\mathbf{x_i}$ to be independent. For computing the DMT, we know from Lemma 2.4 that the inputs can in fact be independent and identically distributed with a $\mathbb{CN}(0, I)$ distribution. So we have

$$
\begin{aligned}
& I(\mathbf{x}; \mathbf{y} | \mathbf{H} = H) \\
& \quad = \sum_{i=1}^{M} I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i)
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\{I(\mathbf{x}; \mathbf{y} | \mathbf{H} = H) \le r \log \rho\} \\
& \quad = \Pr\{\sum_{i=1}^{M} I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i) \le r \log \rho\} \\
& \quad = \Pr\{\sum_{i=1}^{N} \ n_i I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i) \le r \log \rho\}
\end{aligned}
$$

Now, define $Z_i := n_i I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i)$. Also let

$$
\begin{aligned}
\rho^{-\tilde{d}_i(r)} &\doteq \Pr\{Z_i < r \log(\rho)\} \\
&= \Pr\{I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i) < \left(\frac{r}{n_i}\right) \log(\rho)\} \\
&= \rho^{-d_i\left(\frac{r}{n_i}\right)} \\
\text{where, } \rho^{-d_i(r)} &\doteq \Pr\{I(\mathbf{x}_i; \mathbf{y}_i | \mathbf{H}_i = H_i) < r \log(\rho)\}.
\end{aligned}
$$

Using the same convolution argument in the proof of Lemma 2.5,

$$
\begin{aligned}
d(r) &= \inf_{(r_1, r_2, \cdots, r_N):\ \sum_{i=1}^{N} r_i = r} \ \sum_{i=1}^{N} \tilde{d}_i(r) \\
&= \inf_{(r_1, r_2, \cdots, r_N):\ \sum_{i=1}^{N} r_i = r} \ \sum_{i=1}^{N} d_i\left(\frac{r_i}{n_i}\right) \\
&= \inf_{(r_1, r_2, \cdots, r_N):\ \sum_{i=1}^{N} \ n_i r_i = r} \ \sum_{i=1}^{N} d_i(r_i).
\end{aligned}
$$

Acknowledgment

Thanks are due to K. Vinodh and M. Anand for useful discussions.

References

[1] K. Sreeram, S. Birenjith, and P. V. Kumar, "DMT of multi-hop cooperative networks - Part II: Half-duplex networks with full-duplex performance," submitted to *IEEE Trans. Inform. Theory*.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity–Part I: System description," *IEEE Trans. Commun.*, vol. 51, no.11, pp. 1927–1938, Nov. 2003.

[3] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[4] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[5] J. N. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[6] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity–multiplexing tradeoff in half–duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[7] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space–time signal design," *IEEE Jour. Selected Areas in Commun.*, Vol. 22, No. 6, pp. 1099- 1109, Aug. 2004.

[8] Y. Jing, and B. Hassibi, "Distributed space–time coding in wireless relay networks," *IEEE Trans. Wireless Commun.* , Vol. 5, No. 12, pp. 3524-3536, Dec. 2006.

[9] M. Yuksel and E. Erkip, "Multiple–antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective", *IEEE Trans. Inform. Theory* , vol 53, no.10, pp. 3371-3393, Oct. 2007.

[10] P. Elia, K. Vinodh, M. Anand, and P. V. Kumar, "D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols," *IEEE Trans. Inform. Theory,* accepted for publication pending revision. Available Online: http://arxiv.org/abs/cs/0611156 , Nov. 2006.

[11] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: a 2 x 2 full-rate space-time code with non-vanishing determinants," *IEEE Trans. Inform. Theory*, Vol. 51, No. 4, pp. 1432-1436, Apr. 2005.

[12] N. Prasad and M. K. Varanasi, "High performance static and dynamic cooperative communication protocols for the half duplex fading relay channel," *Proc. IEEE Globecom,* Nov. 2006.

[13] S. Tavildar and P. Viswanath, "Approximately universal codes over slow-fading channels" *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3233–3258, July 2006.

[14] P. Elia, K. Raj Kumar, S. A. Pawar, P. V. Kumar, and H-F. Lu, "Explicit, minimum-delay space-time codes achieving the diversity-multiplexing gain tradeoff," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3869–3884, Sept. 2006.

[15] P. Elia and P. V. Kumar, "Approximately–universal space–time codes for the parallel, multi–block and cooperative dynamic–decode–and–forward channels," *Available Online: http://arxiv.org/abs/0706.3502* , June 2007.

[16] K. R. Kumar and G. Caire, "Coding and decoding for the dynamic decode and forward relay protocol, " submitted to *IEEE Trans. on Inform. Theory*, Available Online: http://arxiv.org/abs/0801.2588 , Jan. 2008.

[17] S. Yang and J.-C. Belfiore, "Optimal space–time codes for the MIMO amplify-and-forward cooperative channel," *IEEE Trans. Inform. Theory,* vol. 53, Issue 2, pp 647-663, Feb. 2007.

[18] S. Yang and J.-C. Belfiore, "Diversity of MIMO multihop relay channels," submitted to *IEEE Trans. on Inform. Theory*, Available Online: http://arxiv.org/abs/0708.0386, Aug. 2007.

[19] S. Yang and J.-C. Belfiore, "Towards the optimal amplify–and–forward cooperative diversity scheme," *IEEE Trans. Inform. Theory*, vol. 53, Issue 9, pp 3114-3126, Sept. 2007.

[20] B. A. Sethuraman, B. Sundar Rajan, and V. Shashidhar, "Full–diversity, high–rate, space–time block codes from division algebras," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[21] K. Sreeram, S. Birenjith, and P. V. Kumar, "Multi–hop cooperative wireless networks: Diversity multiplexing tradeoff and optimal code design," *Proceedings of Information Theory and Applications Workshop, UCSD*, Feb. 2008.

[22] K. Sreeram, S. Birenjith, K. Vinod, M. Anand, and P. V. Kumar "On the throughput, DMT and optimal code construction of the K-parallel-path cooperative wireless fading network," *Proc. 10th Int. Symp. Wireless Personal Multimedia Commun.,* Dec. 2007.

[23] K. Sreeram, S. Birenjith, and P. V. Kumar, "DMT of multi–hop cooperative networks–Part I: K–parallel–path–networks" *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, July 6-11, 2008.

[24] K. Sreeram, S. Birenjith, and P. V. Kumar, "Diversity and degrees of freedom of cooperative wireless networks," *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, July 6-11, 2008.

[25] K. Sreeram, S. Birenjith, and P. V. Kumar, "Multi–hop cooperative wireless networks: Diversity multiplexing tradeoff and optimal code design," *Available Online* : http://arxiv.org/pdf/0802.1888 , Feb. 2008.

[26] K. Sreeram, S. Birenjith, and P. V. Kumar, "On the throughput, DMT and optimal code construction of the K–parallel–path cooperative wireless fading network," *USC CSI Technical Report*, CSI-2007-06-07, June 2007.

[27] S. Borade, L. Zheng, and R. Gallager, "Amplify and forward in wireless relay networks: Rate, diversity and network size," *IEEE Trans. Inform. Theory*, vol 53, no.10, pp 3302-3318, Oct. 2007.

[28] C. Rao and B. Hassibi, "Diversity-multiplexing gain trade-off of a MIMO system with relays," *Proc. IEEE Inform. Theory Workshop*, Norway, July 2007.

[29] C. Rao, "Asymptotic analysis of wireless systems with rayleigh fading," *Phd Thesis, California Institute of Technology*, Mar. 2007.

[30] S. O. Gharan, A. Bayesteh, and A. K. Khandani, "On the diversity-multiplexing tradeoff in multiple-relay network," submitted to *IEEE Trans. Inform. Theory*, April 2008.

[31] A. S. Avestimehr, S. N. Diggavi, and D. Tse, "Wireless network information flow," *Proc. Forty-fifth Allerton Conf. Commun. Contr. Comput.*, Illinois, Sep 2007.

[32] A. S. Avestimehr, S. N. Diggavi, and D. Tse, "A deterministic approach to wireless relay networks," *Proc. Forty-fifth Allerton Conf. Commun. Contr. Comput.*, Illinois, Sep 2007.

[33] A. S. Avestimehr, S. N. Diggavi, and D. Tse, "Approximate capacity of gaussian relay networks ," *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, July 6-11, 2008.

[34] B. Nazer and M. Gastpar, " The Case for Structured Random Codes in Network Capacity Theorems," *European Trans. Telecom., Special Issue on New Directions in Inform. Theory,* to appear Spring 2008.

[35] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp.3425-3441, Aug. 2008.

[36] S. A. Jafar and S. Shamai, "Degrees of freedom region for the MIMO X channel," *IEEE Trans. Inform. Theory*, Vol. 54, No. 1, pp:151-170, Jan. 2008.

[37] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over mimo x channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp.3457-3470, Aug. 2008.

[38] V. R. Cadambe, S. A. Jafar, and S. Shamai, "Interference alignment on the deterministic channel and application to fully connected AWGN interference networks," *Available Online*: http://arxiv.org/abs/0711.2547 , Nov. 2007.

[39] C. E. Shannon, "The zero-error capacity of a noisy channel," *IEEE Trans. Inform. Theory*, vol. 2, no. 3, pp. 8–19, Sep 1956.

[40] F. Oggier and B. Hassibi, "Code design for multihop wireless relay networks," accepted for publication in *EURASIP Journal on Advances in Signal Processing* . Available Online: http://www.hindawi.com/journals/asp/raa.457307.html, Nov. 2007.

[41] R. Vaze and R. W. Heath Jr., "Maximizing reliability in multi-hop wireless networks with cascaded space-time codes ," *Proc. Inform. Theory and Appln. Workshop, UCSD*, Feb. 2008.

[42] R. Vaze and R. W. Heath Jr., "To code or not to code in multi-hop relay channels ," *Available Online*: http://arxiv.org/abs/0805.3164

[43] Hsiao-feng Lu, "Explicit construction of multi-block space-time codes that acheive the diversity-multiplexing gain tradeoff," *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, USA , 2006 .

[44] S. Yang, J.-C. Belfiore and G. Rekaya, "Perfect space-time block codes for parallel MIMO channels," *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, USA , 2006.

[45] M. Godavarti, and A. O. Hero III, "Diversity and degrees of freedom in wireless communications,"*Proc. IEEE ICASSP*, May 2002, vol. 3, pp. 2861–2854.

[46] A. Host-Madsen and A. Nosratinia, "The multiplexing gain of wireless networks," *Proc. IEEE Int. Symp. Inform. Theory*, Adeliade, 2005.

[47] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd edition, Springer-Verlag, New York, 1998.

[48] A. W. Marshall and I. Olkin, *Inequalities: Theory of majorization and its applications*, Academic Press, New York, 1979.

[49] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd Edition, John Wiley and Sons, New York, 2006.

[50] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, New York, 1985.

[51] D. B. West, *Introduction to Graph Theory*, Prentice Hall, New Jersey, 2000.

[52] A. R. Amir-Moez, "Extreme properties of eigenvalues of a hermitian transformation and singular values of the sum and product of linear transformations," *Duke Math. J.,* Volume 23, Number 3, 1956, pp 463-476.

[53] J. Boyer, D. D. Falconer, and H. Yanikomeroglu, "Diversity order bounds for wireless relay networks," *Proc. of Wireless Commun. and Networking Conf.,* Mar. 2007.

[54] S. Wei, “Diversity multiplexing tradeoff of asynchronous cooperative diversity in wireless networks, ” *IEEE Trans. Inform. Theory,* Vol. 53, No. 11, pp. 4150-4172, Nov. 2007.

[55] Z. Li and X.-G. Xia, “A simple alamouti space-time transmission scheme for asynchronous cooperative systems,” to appear in *IEEE Signal Processing Letters*, vol. 15, No. 1, pp. 804-807, Jan. 2008.

[56] G. S. Rajan and B. S. Rajan, “Multi-group ML decodable collocated and distributed space time block codes,” submitted to *IEEE Trans. Inform. Theory*. Available Online: http://arxiv.org/abs/0712.2384 , Dec. 2007.

[57] Y. Jing and H. Jafarkhani, “Using orthogonal and quasi-orthogonal designs in wireless relay networks,” *Proc. IEEE Globecom,* Dec. 2006.

[58] Z. Yi and Il-Min Kim, “Single-symbol ML decodable distributed STBCs for cooperative networks,” *IEEE Trans. Inform. Theory,* Vol. 53, No. 8, pp. 2977-2985, Aug. 2007.